



Wilson, Gary (1996) Vibrational raman optical activity of peptides and proteins. PhD thesis

<http://theses.gla.ac.uk/6144/>

Copyright and moral rights for this thesis are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

# **Vibrational Raman Optical Activity of Peptides and Proteins**

**By Gary Wilson**

A Thesis presented in partial fulfilment for the degree  
of Doctor of Philosophy in the Faculty of Science in  
Glasgow University

Chemistry Department  
July 1996  
© G. Wilson



## IMAGING SERVICES NORTH

Boston Spa, Wetherby

West Yorkshire, LS23 7BQ

[www.bl.uk](http://www.bl.uk)

PAGE NUMBERING AS  
ORIGINAL

## **Acknowledgements**

I would like to thank my supervisor Professor Laurence Barron for his advice and encouragement throughout the project.

I am also indebted to the help given to me from other members of my research group, namely Dr. Lutz Hecht, Dr. Alasdair Bell and Dr. Steven Ford.

Finally, I would like to thank the Engineering and Physical Science Research Council for the award of a Research Studentship.



# Contents

Abstract	1
Chapter 1 Introduction	2
Chapter 2 Basic Theory of ROA	8
Chapter 3 Instrumentation	27
Chapter 4 Protein Structure Determination	41
Chapter 5 ROA of Peptides and Polypeptides	49
Chapter 6 A Survey of Protein ROA	78
Chapter 7 Protein Dynamics	130
Appendix ROA Signature Maps	179
References	181
Publications	198

## Abstract

Vibrational Raman optical activity (ROA), which is the difference in the Raman scattering of left and right circularly polarized incident light, has recently emerged as a new and incisive probe of biomolecular structure. This thesis is based on new applications of ROA to some current biochemical problems.

The first chapter is a brief explanation of the origin of chirality and the development of vibrational optical activity with special emphasis on ROA. Chapter 2 is a theoretical analysis of ROA and provides a fundamental explanation of the phenomenon. This involves a description as to how the ROA effect is generated using molecular property tensors. The third chapter concentrates on the instrumentation required to measure ROA and the importance of CCD detectors and holographic notch filters in establishing the technique with respect to biopolymers.

Chapter 4 is a brief introduction to protein structure and includes an analysis of the strengths and weaknesses of current biophysical techniques used for structure determination.

Chapters 5 and 6 describe detailed applications of ROA to polypeptides and native proteins. The polypeptides are a suitable starting point since from other spectroscopic techniques they are known to adopt certain conformations, such as  $\alpha$ -helix,  $\beta$ -sheet and random coil. Native proteins are examined in Chapter 6 and the ability of ROA to detect not only secondary but also tertiary structure is highlighted. Chapter 7 is concerned with the important topics of the structure and dynamics of unfolded proteins, molten globules and ligand bound proteins. Finally, in the appendix there is a summary of the assignments made to secondary structure and to loops and turns.

# **Chapter 1**

## **Introduction**

This initial chapter will introduce the phenomena of optical activity and the concept of chirality. It will also provide a brief account of chiroptical spectroscopy and how the new technique of vibrational optical activity has developed.

### **1.1 Optical Rotation**

Viewing the colours of sunlight as it passed along the optical axis of a quartz crystal which was placed between crossed polarizers, Arago made the first observation of optical activity in 1811 (Arago, 1811). In the following year Biot (1812) deduced that the source of the colours depended on the rotation of the plane of polarization of linearly polarized light (i.e. optical rotation) and the fact that the amount of rotation was dependent on the wavelength of the light (i.e. optical rotatory dispersion).

Biot later discovered a second form of quartz which rotated the plane of polarized light in the opposite sense. Biot was also the first to observe optical rotation in organic liquids (Biot, 1815) such as turpentine, camphor and aqueous solutions of sugar and tartaric acid. Even at this time it was beginning to be understood that the origin of the optical activity in fluids was due to a helical structure of the individual molecules, but that in the case of the quartz it was a property of the crystal structure, since molten quartz is not optically active.

In 1824 Fresnel discovered right and left circularly polarized light which allowed him to define optical rotation in terms of different refractive indices for the coherent right and left circularly polarized components into which a linearly polarized light beam can be resolved (Figure 1.1).

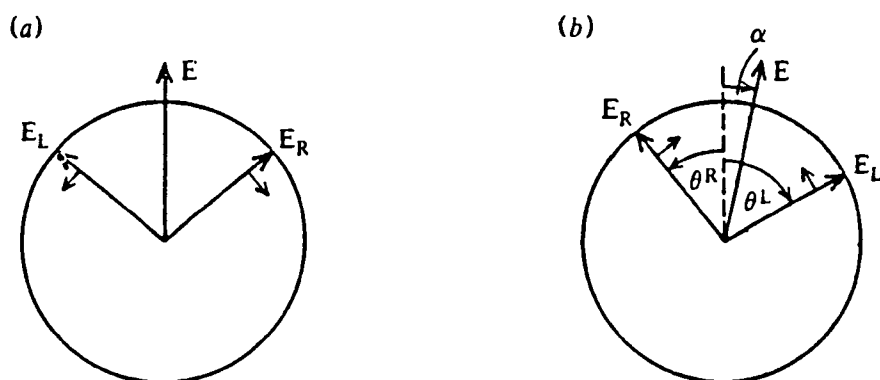


Figure 1.1 (a) The electric field vector decomposed into coherent right and left circularly polarized components.  
(b) The rotated electric field vector has undergone a rotation of  $\alpha$  in the optically active medium.

When, in 1848, Pasteur resolved a racemic mixture of tartaric acid into its two mirror-image enantiomorphous forms it finally proved Fresnel's theory that individual optically active molecules must have an intrinsic handedness. This observation emphasized that molecules are 3-dimensional and therefore stereochemistry was introduced into science.

An interesting observation made by Faraday in 1846 occurred when he induced optical activity in an optically inactive sample by applying a magnetic field. He observed optical rotation in a rod of lead borate glass placed between the poles of an electromagnet with holes bored through the pole pieces to allow a linearly polarized light beam to pass through. This is known as the Faraday effect and implies that electromagnetism and light are intimately linked.

It was not until 1884 that the now commonly used word chirality was introduced by Glasgow's Lord Kelvin. Lord Kelvin used the word *chiral* to describe a geometrical figure "if its image in a plane mirror, ideally realized, cannot be brought into coincidence with itself". The topic of chirality is deep and subtle and is important in physics, chemistry and biology. Recently, for example, the unification

of the weak and electromagnetic forces has revealed that the parity violation present in the weak interactions is present to a tiny extent in all electromagnetic phenomena, which has led to the discovery that enantiomers of a chiral molecule have a very small parity-violating energy difference (MacDermott & Tranter, 1989).

## 1.2 Circular Dichroism Spectroscopy

When polarized incident light interacts with a chiral molecule at an electronic transition frequency there is a differential absorption of the left and right circularly polarized light and this effect is known as electronic circular dichroism (ECD). Proteins have chiral structures and have been the topic of many extensive ECD studies (see later).

The optical rotatory dispersion (ORD), together with the ellipticity of the rotated beam due to CD, in the region of an electronic transition frequency is called the Cotton effect, in honour of the French physicist Aimé Cotton, who performed the pioneering studies of the wavelength-dependent aspects of these phenomena (Cotton, 1895).

## 1.3 Vibrational Optical Activity

Natural vibrational optical activity (VOA) has developed over the past 20 years into a powerful new form of chiroptical spectroscopy (Barron, 1982; Nafie *et al.*, 1994). The driving force behind this has been the realization that the measurement of optical activity associated with molecular vibrations would provide a wealth of new stereochemical information. There have been two streams of development in vibrational spectroscopy; namely the infrared (IR) absorption form and the Raman scattering form. The IR form of VOA is called vibrational circular

dichroism (VCD) and is a direct extension of ECD into the IR region of the spectrum (Stephens & Lowe, 1985; Freedman & Nafie, 1994). VCD is therefore the differential absorption of left and right circularly polarized IR radiation in a vibrational absorption band. The complementary Raman technique, called Raman optical activity (ROA), depends on the differential Raman scattering of right and left circularly polarized visible incident light for a corresponding vibrational Raman scattering transition (Barron, 1982).

The first measurements of ROA (Barron *et al.*, 1973a,b,c) and VCD (Holzwarth *et al.*, 1974; Nafie *et al.*, 1975), were performed in the early 1970's. It quickly became apparent that due to the weakness of Raman scattering compared to IR absorption, VCD was much easier to measure experimentally than ROA (there is also a greater problem with artifacts in ROA which will be discussed in Chapter 3). By the 1980's VCD had developed into a routine technique in several laboratories to solve stereochemical and biochemical problems. *Ab initio* computations were also developed to interpret VCD spectra (Lowe *et al.*, 1986). Fourier-transform methods were also performed at the early stages of VCD development (Nafie *et al.*, 1979). During this time ROA was left behind, with interpretation based on correlations and simple models.

Recently, though, advances in optical and electronic technology have propelled ROA from being just an interesting phenomenon into a powerful and incisive new probe of stereochemistry. This has mainly involved the use of charge-coupled device (CCD) detectors, holographic notch filters and the dramatic increase in the signal-to-noise ratio by using a backscattering configuration (Barron *et al.*, 1989; Hecht *et al.*, 1991), which will be discussed in greater detail later (Hecht *et al.*, 1989; Che *et al.*, 1991). An *ab initio* theoretical model has also now been developed for ROA which greatly helps the interpretation of ROA spectra (Polavarapu, 1990).

The one photon origin of VCD is illustrated in Figure 1.2, which shows a vibrational transition between sublevels  $g_0$  and  $g_1$  in the ground electronic state.

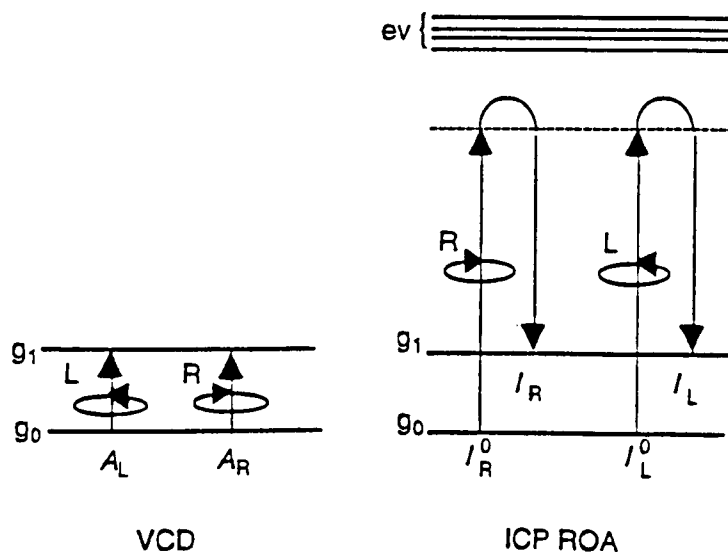


Figure 1.2 Energy-level polarization diagrams for VCD and ICP ROA

In the second diagram in Figure 1.2 the two photon origin of ROA is illustrated. The excited vibronic states (ev) can be far-from-resonance, pre-resonant or in strong resonance with respect to the incident or Raman scattered photons. The current Glasgow set-up involves a far-from-resonance scattering technique.

VCD can be described as the difference in absorbance of a sample

$$\Delta A = A_L - A_R \quad (1.1)$$

where  $A_L$  and  $A_R$  describe the absorption of left and right circularly polarized incident light. The ROA form can be expressed as the circular intensity difference (CID) of the Raman scattered light

$$\Delta I = I^R - I^L \quad (1.2)$$

where  $I^R$  and  $I^L$  are the Raman scattered intensities in right and left circularly polarized light.

The great advantage of vibrational optical activity is that, since a vibrational spectrum of a chiral molecule can contain up to  $3N-6$  normal modes of vibration, each of which can show optical activity, a vibrational optical activity spectrum will

contain information about the conformation and absolute configuration of the entire structure. This is in contrast to an ECD spectrum of a chiral molecule that will contain a single vibronic band due to for example the  $\pi^* \leftarrow n$  transition of the carbonyl chromophore. ECD will therefore just probe the chiral environment around the chromophore whereas an ROA or VCD spectrum will contain bands that are associated with every part of the molecule.

ROA is a currently emerging chiroptical technique which appears to have a higher level of sensitivity towards structural features of biopolymers than ECD or VCD and will be the subject of this thesis, which will focus mainly on peptides and proteins.



## Chapter 2

### Basic Theory of ROA

In this chapter there will be a discussion of the ROA effect, with simple models used to explain how it arises. The ROA phenomenon is a form of Raman spectroscopy which has been explained in great detail elsewhere (Long, 1977; Carey, 1982). In particular, the bond polarizability model will be highlighted since this exposes the vital importance of a backscattering geometry for biopolymer studies.

#### 2.1 Natural Raman Optical Activity (ROA)

The ROA measurement is associated with a vibrational transition of angular frequency  $\omega_v$ , where  $I^R$  and  $I^L$  are the Raman scattered intensities at the angular frequency  $(\omega - \omega_v)$  for the right and left circularly polarized incident light of frequency  $\omega$ , as shown in Figure 2.1.

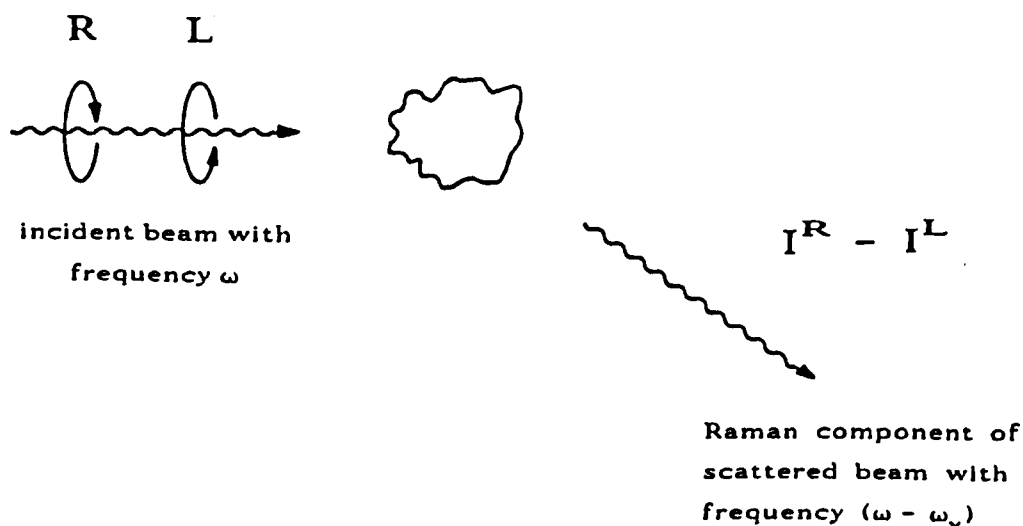


Figure 2.1 The basic ROA measurement

It is important to realize that Rayleigh (no change in frequency of the scattered beam) and Raman optical activity will be observed in all directions with different intensities. A scattering picture for both optical rotation and ROA can be generated using a simple two-group model, illustrated in Figure 2.2, which shows that the two phenomena have quite different origins.

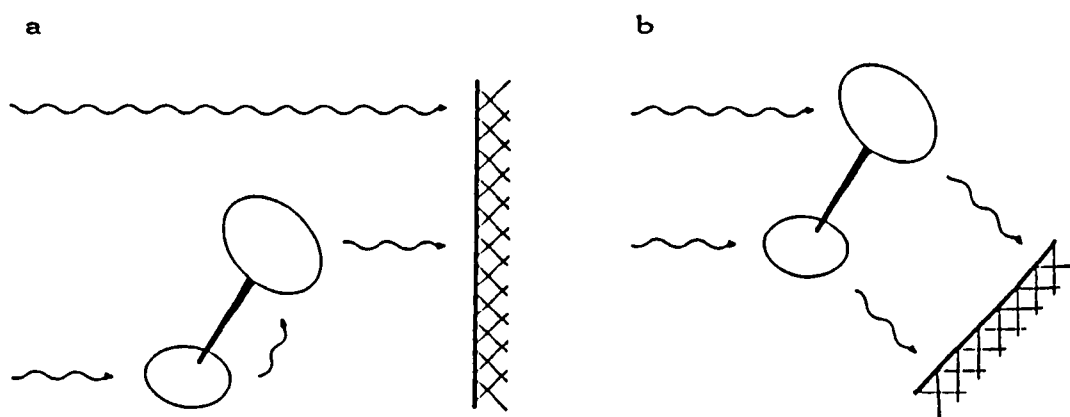


Figure 2.2      A two group model representing  
 (a) Optical rotation  
 (b) Rayleigh and Raman optical activity

The two anisotropic groups are held in a twisted chiral arrangement and therefore represent a chiral species. Natural optical rotation as shown in Figure 2.2(a) occurs via the Kirkwood (1937) dynamic coupling mechanism. The photon scattering involves an unscattered photon interfering with a forward-scattered photon at the detector (Barron, 1982). It is only the second photon that has sampled the chirality of the molecule by being deflected from one of the anisotropic groups to the other and then interacting with the first photon in the forward direction. This is in contrast to the ROA effect in Figure 2.2(b) where both photons interact independently with the chiral molecule. No coupling is required between the two-groups (although this does

provide higher-order contributions) because the interference between these two photons at the detector is enough to probe the chirality and conformation of the molecule (Barron & Buckingham, 1974).

## 2.2 The Polarizability and Optical Activity Tensors

The fundamental scattering mechanism for ROA was discovered in the late 1960's (Atkins & Barron, 1969; Barron, 1969) and is based on the interference between light waves scattered *via* the polarizability and optical activity tensors which will be explained shortly. This initial theory was based on a quantum field development. A later version using semiclassical methods was developed as the definitive theory of ROA by Barron and Buckingham (1971). This involved calculating the electric field vector radiated by the oscillating electric dipole, magnetic dipole and electric quadrupole moment induced by the right and left circularly polarized radiation. This then allowed the intensities for each of these interactions to be calculated. Induced molecular electric multipole moments are set into oscillation by the light wave when it encounters the molecule. The secondary waves created by these oscillating multipole moments are scattered in all directions.

Cartesian tensor notation will be used extensively in these equations and therefore needs to be explained (Jeffreys, 1931; Barron, 1982). A Greek subscript denotes a vector or tensor component and can be equal to  $X, Y$  or  $Z$ . A repeated Greek suffix means summation over all three components (e.g.  $a_\alpha b_\alpha \equiv \mathbf{a} \cdot \mathbf{b} = a_x b_x + a_y b_y + a_z b_z$ ).

Therefore using this definition, the electric dipole, magnetic dipole and traceless electric quadrupole defined in SI units are (Barron, 1982)

$$\mu_\alpha = \sum_i e_i r_{i\alpha} \quad (2.1a)$$

$$m_a = \sum_i (e_i / 2m_i) \epsilon_{\alpha\beta\gamma} r_{i\beta} p_{i\gamma} \quad (2.1b)$$

$$\Theta_{\alpha\beta} = \frac{1}{2} \sum_i e_i (3r_{i\alpha} r_{i\beta} - r_i^2 \delta_{\alpha\beta}) \quad (2.1c)$$

Particle  $i$  is denoted as having position vector  $\mathbf{r}_i$ , electric charge  $e_i$ , mass  $m_i$  and linear momentum  $\mathbf{p}_i$ . The symbol  $\delta_{\alpha\beta}$  is the unit second-rank symmetric tensor and  $\epsilon_{\alpha\beta\gamma}$  represents the unit third-rank antisymmetric (Levi-Civita) tensor. The  $\alpha$  component of the vector product  $\mathbf{r} \times \mathbf{p}$  is thus represented by  $\epsilon_{\alpha\beta\gamma} r_\beta p_\gamma$ .

Light is an electromagnetic phenomenon and can be represented by an electric field vector of angular frequency  $\omega = 2\pi c / \lambda$ . For a plane-wave travelling in the direction of the unit vector  $\mathbf{n}$  at velocity  $c$

$$\mathbf{E} = \mathbf{E}^{(0)} \exp [-i\omega(t - \mathbf{n} \cdot \mathbf{r}/c)] \quad (2.2)$$

$\mathbf{E}^{(0)}$  represents the amplitude of the wave.

This electric vector induces a real oscillating electric dipole, magnetic dipole and electric quadrupole moment in the molecule

$$\mu_\alpha = \alpha_{\alpha\beta} E_\beta + (1/\omega) G'_{\alpha\beta} \dot{B}_\beta + \frac{1}{3} A_{\alpha\beta\gamma} \nabla_\beta E_\gamma + \dots \quad (2.3a)$$

$$m_\alpha = -(1/\omega) G'_{\alpha\beta} \dot{E}_\beta + \dots \quad (2.3b)$$

$$\Theta_{\alpha\beta} = A_{\alpha\beta\gamma} E_\gamma + \dots \quad (2.3c)$$

These fields and field gradients are specified at the molecular origin used to calculate the molecular multipole moments. Quantum-mechanical expressions for the real dynamic molecular property tensors obtained from time dependent perturbation theory are as follows

$$\alpha_{\alpha\beta} = \frac{2}{\hbar} \sum_{j \neq n} (\omega_{jn} / \omega_{jn}^2 - \omega^2) \text{Re}(\langle n | \mu_\alpha | j \rangle \langle j | \mu_\beta | n \rangle) \quad (2.4a)$$

$$G'_{\alpha\beta} = -\frac{2}{\hbar} \sum_{j \neq n} (\omega / \omega_{jn}^2 - \omega^2) \text{Im}(\langle n | \mu_\alpha | j \rangle \langle j | m_\beta | n \rangle) \quad (2.4b)$$

$$A_{\alpha\beta\gamma} = \frac{2}{\hbar} \sum_{j \neq n} (\omega_{jn} / \omega_{jn}^2 - \omega^2) \text{Re}(\langle n | \mu_\alpha | j \rangle \langle j | \Theta_{\beta\gamma} | n \rangle) \quad (2.4c)$$

In these equations  $n$  and  $j$  represent the initial state and the virtual intermediate state involved in the Raman transition, and  $\omega_{jn}$  defines the angular frequency separation of states  $n$  and  $j$ . Equation 2.4a is the electric dipole-electric dipole tensor where  $\alpha_{\alpha\beta}$  is the polarizability that is usually associated with refraction, light scattering, the van der Waals force etc. The electric dipole-magnetic dipole optical activity tensor (equation 2.4b) is represented by  $G'_{\alpha\beta}$  and its trace  $G'_{\alpha\alpha}$  is responsible for conventional optical rotation in fluids.  $A_{\alpha\beta\gamma}$  (equation 2.4c) is the electric dipole-electric quadrupole optical activity tensor which can make significant contributions to the optical rotation in oriented samples. From these three basic expressions the ROA phenomenon can be described (Barron, 1982).

### 2.3 Optical Rotation in Isotropic Samples

It is interesting to compare the expressions used to calculate ROA and those that are used to generate conventional optical rotation. In an isotropic (randomly orientated) sample the natural optical rotation angle is given by the Rosenfield equation (1928)

$$\Delta\theta = -\frac{1}{3} \omega \mu_o l N G'_{\alpha\alpha} = -\frac{1}{3} \omega \mu_o l N (G'_{xx} + G'_{yy} + G'_{zz}) \quad (2.5)$$

where  $l$  is the finite path length of the sample,  $N$  is the number of chiral molecules per unit volume and  $\mu_o$  is the vacuum permeability constant.

For a crystal with oriented chiral molecules and for the light beam

propagating in the z direction the optical rotation is given by the Buckingham-Dunn equation (1971)

$$\Delta\theta = -\frac{1}{2} \omega \mu_o l N \left[ \frac{1}{3} \omega (A_{xyz} - A_{yxz}) + G'_{xx} + G'_{yy} \right] \quad (2.6)$$

It is only in oriented samples that there are electric dipole-electric quadrupole contributions to natural optical rotation and CD. In isotropic samples the electric dipole-electric quadrupole contributions in optical rotation and CD average to zero. In ROA it is important to realize that these contributions are retained in isotropic situations and make a contribution that is of the same order as the electric dipole-magnetic dipole tensor. This is therefore an additional factor to the optical activity of ROA.

## 2.4 The ROA Observables

The dominant contributions to Rayleigh and Raman scattering is dependent on  $\alpha^2$  and the ROA is produced by the combination of  $\alpha G'$  and  $\alpha A$  which have opposite signs in right and left circularly polarized incident light. ROA is usually quantified as the dimensionless CID (Barron & Buckingham, 1971)

$$\Delta_{ICP} = \frac{I^R - I^L}{I^R + I^L} \quad (2.7)$$

where  $I^R$  and  $I^L$  are the Raman scattered intensities for right and left circularly polarized incident light.

ROA can be measured using three different polarization modulations. The incident circular polarization (ICP) approach is adopted at Glasgow and depends on the differential scattering of right and left circularly polarized incident light (Hecht *et al.*, 1992). The second form is called scattered circular polarization (SCP) ROA (Barron & Buckingham, 1975; Spencer *et al.*, 1988) and is defined as the excess of

right or left circularly polarized intensity in the Raman scattered light and referred to as the degree of circularity. This SCP ROA can be expressed in dimensionless form

$$\Delta_{SCP} = \frac{I_R - I_L}{I_R + I_L} \quad (2.8)$$

where  $I_R$  and  $I_L$  represent the excess of right and left circularly polarized intensity in the Raman scattered light from a linear polarized incident beam.

After SCP ROA was detected in 1988 it was realized that two other forms of ROA existed and that they could be measured using a simultaneous combination of the techniques employed in ICP and SCP ROA (Nafie & Freedman, 1989). These two new forms were called dual circular polarization (DCP) ROA and if measured when the circular polarization modulation of the incident and scattered beams are in-phase it is called DCP<sub>I</sub> ROA; or if the incident and scattered beams are out-of-phase it is called DCP<sub>II</sub> ROA. The CID's can be expressed as follow

$$\Delta DCP_I = \frac{I_R^R - I_L^L}{I_R^R + I_L^L} \quad (2.9a)$$

$$\Delta DCP_{II} = \frac{I_L^R - I_R^L}{I_L^R + I_R^L} \quad (2.9b)$$

where  $I_L^R$  for example represents the intensity of left polarized scattered radiation generated from right circularly polarized incident radiation.

We shall now focus on ICP ROA since that is the current configuration of the Glasgow instrument.

The scattering angle also defines different ROA experimental configurations. The usual scattering directions are forward (0°), right-angle (90°) and backward (180°). In right-angle scattering, a linear polaroid analyser placed either perpendicular (x) or parallel (z) to the scattering plane (yz) allows two different measurements to be made. These give respectively polarized and depolarized ROA

and are analogous to the polarized and depolarized Raman intensities which are used to calculate the depolarization ratio.

In terms of molecular property tensors, the ROA CID expressions for different experimental configurations are found to be (Barron, 1982)

$$\Delta(0^\circ) = \frac{8[45\alpha G' + \beta(G')^2 - \beta(A)^2]}{2c[45\alpha^2 + 7\beta(\alpha)^2]} \quad (2.10a)$$

$$\Delta(180^\circ) = \frac{48[\beta(G')^2 + \frac{1}{3}\beta(A)^2]}{2c[45\alpha^2 + 7\beta(\alpha)^2]} \quad (2.10b)$$

$$\Delta_x(90^\circ) = \frac{2[45\alpha G' + 7\beta(G')^2 + \beta(A)^2]}{c[45\alpha^2 + 7\beta(\alpha)^2]} \quad (2.10c)$$

$$\Delta_z(90^\circ) = \frac{12[\beta(G')^2 - \frac{1}{3}\beta(A)^2]}{6c\beta(\alpha)^2} \quad (2.10d)$$

These expressions are calculated as explained previously and refer specifically to isotropic samples such as liquids and solutions. This means that the polarizability-polarizability and polarizability-optical activity tensor products have been averaged over all possible orientations for the scattering molecule. The tensor component products produced in the averaging process in equations 2.10a - 2.10d are invariant to axis rotations and are thus all that survive in isotropic samples.

The isotropic invariants of the polarizability tensor and the electric dipole-magnetic dipole optical activity tensor are defined as

$$\alpha = \frac{1}{3}\alpha_{aa} = \frac{1}{3}(\alpha_{xx} + \alpha_{yy} + \alpha_{zz}) \quad (2.11a)$$

$$G' = \frac{1}{3}G'_{aa} = \frac{1}{3}(G'_{xx} + G'_{yy} + G'_{zz}) \quad (2.11b)$$



and the anisotropic invariants for the polarizability-polarizability and polarizability-optical activity tensor component products  $\beta(\alpha)^2$ ,  $\beta(G')^2$  and  $\beta(A)^2$  as

$$\begin{aligned}\beta(\alpha)^2 &= \frac{1}{2} (3\alpha_{\alpha\beta} \alpha_{\alpha\beta} - \alpha_{\alpha\alpha} \alpha_{\beta\beta}) \\ &= \frac{1}{2} [(\alpha_{xx} - \alpha_{yy})^2 + (\alpha_{xx} - \alpha_{zz})^2 + (\alpha_{yy} - \alpha_{zz})^2 + 6(\alpha_{xy}^2 + \alpha_{xz}^2 + \alpha_{yz}^2)]\end{aligned}\quad (2.12a)$$

$$\begin{aligned}\beta(G')^2 &= \frac{1}{2} (3\alpha_{\alpha\beta} G'_{\alpha\beta} - \alpha_{\alpha\alpha} G'_{\beta\beta}) \\ &= \frac{1}{2} \{[(\alpha_{xx} - \alpha_{yy})(G'_{xx} - G'_{yy}) + (\alpha_{xx} - \alpha_{zz})(G'_{xx} - G'_{zz}) \\ &\quad + (\alpha_{yy} - \alpha_{zz})(G'_{yy} - G'_{zz})] + 3[\alpha_{xy}(G'_{xy} + G'_{yx}) \\ &\quad + \alpha_{xz}(G'_{xz} + G'_{zx}) + \alpha_{yz}(G'_{yz} + G'_{zy})]\}\end{aligned}\quad (2.12b)$$

$$\begin{aligned}\beta(A)^2 &= \frac{1}{2} \omega \alpha_{\alpha\beta} \varepsilon_{\alpha\gamma\delta} A_{\gamma\delta\beta} \\ &= \frac{1}{2} \omega [(\alpha_{yy} - \alpha_{xx}) A_{zxy} + (\alpha_{xx} - \alpha_{zz}) A_{yzx} + (\alpha_{zz} - \alpha_{yy}) A_{xyz} \\ &\quad + \alpha_{xy} (A_{yyz} - A_{zyy} + A_{zxx} - A_{xxz}) \\ &\quad + \alpha_{xz} (A_{yzz} - A_{zzx} + A_{xxy} - A_{yxx}) \\ &\quad + \alpha_{yz} (A_{zzx} - A_{xzz} + A_{xyy} - A_{yyx})]\end{aligned}\quad (2.12c)$$

All these components in these isotropic invariants are specified with respect to molecule-fixed axes. It should be noted that in equations 2.10a - 2.10d the only contributions to the ROA are made by  $\alpha G'$ ,  $\beta(G')^2$  and  $\beta(A)^2$  because only they are pseudoscalars and so depend on the sense of circular polarization. Although terms such as  $G^2$  and  $A^2$  make contributions to the ROA intensity they are scalars and so do not have any dependence on the sense of circular polarization.

All of the CID expressions so far are only valid at transparent frequencies (i.e. far-from-resonance) and are also only a description of Rayleigh scattering and thus need to be extended to cover the vibrational Raman form. This requires the use of transition tensor versions of the polarizability and optical activity tensors. The conversion of the Rayleigh form to the Raman occurs as follows

$$\alpha_{\alpha\beta} \rightarrow \langle m_\nu | \alpha_{\alpha\beta}(Q) | n_\nu \rangle \quad (2.13a)$$

$$G'_{\alpha\beta} \rightarrow \langle m_\nu | G'_{\alpha\beta}(Q) | n_\nu \rangle \quad (2.13b)$$

$$A_{\alpha\beta\gamma} \rightarrow \langle m_\nu | A_{\alpha\beta\gamma}(Q) | n_\nu \rangle \quad (2.13c)$$

The transition involves the initial and final vibrational states,  $n_\nu$  and  $m_\nu$ . The terms  $\alpha_{\alpha\beta}(Q)$ ,  $G'_{\alpha\beta}(Q)$  and  $A_{\alpha\beta\gamma}(Q)$  are the effective electronic polarizability and optical activity operators and they depend on the normal vibrational coordinates  $Q_p$ . Using the Placzek polarizability theory of Raman intensities at transparent frequencies and using a Taylor series, the transition polarizability can be expanded in the normal vibrational coordinates (Long, 1977; Barron, 1982)

$$\begin{aligned} \langle m_\nu | \alpha_{\alpha\beta}(Q) | n_\nu \rangle &= (\alpha_{\alpha\beta})_0 \delta_{m_\nu n_\nu} \\ &+ \sum_p \left( \frac{\partial \alpha_{\alpha\beta}}{\partial Q_p} \right)_0 \langle m_\nu | Q_p | n_\nu \rangle + \dots \end{aligned} \quad (2.14)$$

The first term describes the Rayleigh scattering and the second the vibrational Raman scattering. The subscript zero indicates that the function is evaluated at the equilibrium nuclear configuration. For the Raman part, in the case of a simple harmonic oscillator the selection rule  $\Delta_\nu = \pm 1$  exists.

Using the Taylor expansion in equation 2.14 the following expressions for the Raman intensity and optical activity can be found

$$\langle 0 | \alpha_{\alpha\beta} | 1_p \rangle \langle 1_p | \alpha_{\alpha\beta} | 0 \rangle = \left( \frac{\hbar}{2\omega_p} \right) \left( \frac{\partial \alpha_{\alpha\beta}}{\partial Q_p} \right)_0 \left( \frac{\partial \alpha_{\alpha\beta}}{\partial Q_p} \right)_0 \quad (2.15a)$$

$$\langle 0 | \alpha_{\alpha\beta} | 1_p \rangle \langle 1_p | G'_{\alpha\beta} | 0 \rangle = \left( \frac{\hbar}{2\omega_p} \right) \left( \frac{\partial \alpha_{\alpha\beta}}{\partial Q_p} \right)_0 \left( \frac{\partial G'_{\alpha\beta}}{\partial Q_p} \right)_0 \quad (2.15b)$$

$$\langle 0 | \alpha_{\alpha\beta} | 1_p \rangle \langle 1_p | \epsilon_{\alpha\gamma\delta} A_{\gamma\delta\beta} | 0 \rangle = \left( \frac{\hbar}{2\omega_p} \right) \left( \frac{\partial \alpha_{\alpha\beta}}{\partial Q_p} \right)_0 \epsilon_{\alpha\gamma\delta} \left( \frac{\partial A_{\gamma\delta\beta}}{\partial Q_p} \right)_0 \quad (2.15c)$$

These equations are valid for the fundamental transition  $1_p \leftarrow 0$  associated with the normal vibrational coordinate  $Q_p$ , where 0 represents the ground state and  $1_p$  represents the first excited state of the molecular vibration associated with the normal coordinate  $Q_p$ . It is also helpful to point out that  $\hbar/2\omega_p$  is equal to  $|\langle 1_p | Q_p | 0 \rangle|^2$ .

From an inspection of equations 2.15a - 2.15c, it is clear that the ROA phenomenon originates in the interference between the polarizability and the optical activity tensors of the molecule. This observation also allows the basic symmetry requirements for ROA to be deduced. The symmetry restrictions are thus that the same components of  $\alpha_{\alpha\beta}$  and  $G'_{\alpha\beta}$  must span the irreducible representation of the particular normal mode of vibration. This limits the molecules that can show ROA to those with the chiral point groups  $C_n$ ,  $D_n$ ,  $O$ ,  $T$  and  $I$ . This can only happen if the polar and axial tensors have the same rank, such that  $\alpha_{\alpha\beta}$  transforms like  $\mu_\alpha \mu_\beta$  and  $G'_{\alpha\beta}$  transforms like  $\mu_\alpha m_\beta$ . It should also be noted that in spite of the fact that  $A_{\alpha\beta\gamma}$  does not transform like  $G'_{\alpha\beta}$ , the second-rank axial tensor  $\epsilon_{\alpha\gamma\delta} A_{\gamma\delta\beta}$  in equation 2.15c which combines with  $\alpha_{\alpha\beta}$  does have the same transformation properties as that for  $G'_{\alpha\beta}$ . The result of this is that all Raman active vibrations in a chiral molecule have the possibility of producing ROA.

## 2.5 The Bond Polarizability Theory of ROA

The bond polarizability theory of ROA (Barron, 1982; Escibano & Barron, 1988) allows simple models to be used to understand the generation of ROA. Two different models will now be discussed: the first is a simple two-group model and the second will be a general bond polarizability theory.

### 2.5(a) A Simple Two-Group Model

The origin of ROA in the two-group model is illustrated in Figure 2.2 which

shows that ROA arises from interference at the detector between waves scattered independently by the two separate groups that make up the chiral species. The theoretical treatment of the model can be approached in two different ways. The original development (Barron & Buckingham, 1974) depends on the polarizability tensor  $\alpha_{\alpha\beta}$  being origin independent, whereas the optical activity tensors  $G'_{\alpha\beta}$  and  $A_{\alpha\beta\gamma}$  are dependent on the origin. It follows that if the molecular origin moves from  $\underline{0}$  to  $\underline{0} + \underline{a}$  then (Buckingham & Longuet-Higgins, 1968)

$$\alpha_{\alpha\beta} \rightarrow \alpha_{\alpha\beta} \quad (2.16a)$$

$$G'_{\alpha\beta} \rightarrow G'_{\alpha\beta} + \frac{1}{2} \omega \varepsilon_{\beta\gamma\delta} a_{\gamma} \alpha_{\alpha\delta} \quad (2.16b)$$

$$A_{\alpha\beta\gamma} \rightarrow A_{\alpha\beta\gamma} - \frac{3}{2} a_{\beta} \alpha_{\alpha\gamma} - \frac{3}{2} a_{\gamma} \alpha_{\alpha\beta} + a_{\delta} \alpha_{\alpha\delta} \delta_{\beta\gamma} \quad (2.16c)$$

This initially involves calculating the polarizability and optical activity tensors of the two anisotropic groups referred to local origins. Then using the sum of the local group tensors the local group optical activity tensors are now calculated from a single molecular origin. Applying this method leads to the following isotropic and anisotropic invariants assuming cylindrical bond symmetry

$$\alpha G' = 0 \quad (2.17a)$$

$$\beta(G')^2 = \beta(A)^2 = -\frac{3}{4} \omega \varepsilon_{\beta\gamma\delta} R_{21\gamma} \alpha_{1\alpha\beta} \alpha_{2\delta\alpha} \quad (2.17b)$$

$R_{21}$  represents the vector between the two anisotropic groups. Hence in this model  $\beta(G')^2$  is equivalent to  $\beta(A)^2$  (i.e. the magnetic dipole and electric quadrupole anisotropic contributions are equal), and there is no isotropic ROA .

The second distinct approach to the two-group model does not involve the use of optical activity tensors; it involves the calculation of the electric field amplitudes scattered independently by the two-groups *via* the summing of group

polarizability tensors. Although this is a completely different method to that described previously, the same ROA expressions are obtained at the end for the case where the wavelength of the incident light is much greater than the distance between the two anisotropic groups (Barron, 1982).

An important aspect that has been neglected so far in the discussion is the vibrations of the two-group structure. There will be bond stretches and angle deformations of the two groups that will make contributions to the normal modes  $Q_p$ . These internal vibrational coordinates  $q_i$  will thus alter the interaction between the two groups and a simple example to illustrate this is the coupled oscillator model. If we consider two idealized normal coordinates  $Q_+$  and  $Q_-$  that contain symmetric and antisymmetric combinations of two internal coordinates  $s_1$  and  $s_2$  then

$$Q_+ = N_1 s_1 + N_2 s_2 \quad (2.18a)$$

$$Q_- = N_2 s_1 - N_1 s_2 \quad (2.18b)$$

$N_1$  and  $N_2$  represent constants that are identical for equivalent groups. A coupling mechanism will occur between these two local modes that will lift the degeneracy of  $Q_+$  and  $Q_-$  through some sort of perturbation such as potential energy coupling, kinetic energy coupling and dipolar coupling. This will effectively produce two different Raman bands that will have slightly different frequencies and different intensities. This leads to the fact that the ROA intensities will have equal magnitude but opposite sign, independent of what the two groups are. For the magnetic dipole and electric quadrupole anisotropic ROA invariants it is found that for  $Q_+$  and  $Q_-$

$$[\beta(G')^2]_{\pm} = [\beta(A)^2]_{\pm} = \mp \frac{3}{4} N_1 N_2 \omega \epsilon_{\beta\gamma\delta} R_{21\gamma} \langle 0 | \alpha_{1a\beta} | 1_1 \rangle \langle 1_2 | \alpha_{2ba} | 0 \rangle \quad (2.19)$$

where  $1_i$  represents the first excited state of the local oscillator on group  $i$ . It is apparent that this mechanism will produce adjacent, overlapping bands of equal magnitude but opposite sign. This will lead to the generation of conservative ROA

couplets which as will be observed later are a common feature in ROA spectra.

### 2.5(b) A General Bond Polarizability Theory

The general bond polarizability theory is a hybrid of the two-group model and the inertial model (Barron & Buckingham, 1979). The inertial model differs from the just described two-group model since the ROA mechanism depends on the changing interaction of the radiation field within a chiral species whose framework is continually twisting to cancel out the twist of the torsioning group in order that the corresponding normal mode of vibration generates no resultant angular momentum.

The theory involves the use of the transition polarizability-transition polarizability and transition polarizability-transition optical activity tensors as shown in equations 2.15a - 2.15c. The polarizability for the whole molecule is then written as the sum of the local bond polarizability operators  $\alpha_{i\alpha\beta}$

$$\alpha_{\alpha\beta} = \sum_i \alpha_{i\alpha\beta} \quad (2.20)$$

The transition polarizability can then be calculated from the molecular polarizability with normal vibrational coordinates  $Q_p$  by summing the derivatives of the local internal coordinates  $s_q$  i.e. local bond stretches, angle bends and torsions

$$\langle 1_p | \alpha_{\alpha\beta} | 0 \rangle = \left( \frac{\hbar}{2\omega_p} \right)^{\frac{1}{2}} \sum_i \sum_q \left( \frac{\partial \alpha_{i\alpha\beta}}{\partial s_q} \right)_0 L_{qp} \quad (2.21)$$

$L_{qp}$  represents the elements of the L-matrix allowing  $s_q$  to be defined as linear combinations of  $Q_p$  (from normal coordinate analysis). The optical activity tensors  $G'_{i\alpha\beta}$  and  $A_{i\alpha\beta}$ , referred to a general molecular origin, are then incorporated to provide the following origin-independent expression for ROA (Barron, 1982; Escribano &

Barron, 1988; Barron & Clark, 1982).

$$\begin{aligned}
 \langle 0 | \alpha_{\alpha\beta} | 1_p \rangle \langle 1_p | T_{\alpha\beta} | 0 \rangle = & \\
 - \left( \frac{\hbar\omega}{4\omega_p} \right) \varepsilon_{\beta\gamma\delta} \sum_{i < j} R_{ji\gamma} \left[ \sum_q \left( \frac{\partial \alpha_{i\alpha\beta}}{\partial s_q} \right)_0 L_{qp} \right] \left[ \sum_r \left( \frac{\partial \alpha_{j\delta\alpha}}{\partial s_r} \right)_0 L_{rp} \right] & \\
 - \left( \frac{\hbar\omega}{4\omega_p} \right) \varepsilon_{\beta\gamma\delta} \left[ \sum_i \sum_q \left( \frac{\partial \alpha_{i\alpha\beta}}{\partial s_q} \right)_0 L_{qp} \right] \left[ \sum_j (\alpha_{j\delta\alpha})_0 \sum_r \left( \frac{\partial r_{j\gamma}}{\partial s_r} \right)_0 L_{rp} \right] & \\
 + \left( \frac{\hbar}{2\omega_p} \right) \left[ \sum_i \sum_q \left( \frac{\partial \alpha_{i\alpha\beta}}{\partial s_q} \right)_0 L_{qp} \right] \left[ \sum_j \sum_r \left( \frac{\partial T_{\alpha\beta}}{\partial s_r} \right)_0 L_{rp} \right] & \quad (2.22)
 \end{aligned}$$

$T_{\alpha\beta}$  in the above expression is equal to  $G'_{\alpha\beta}$  or  $1/3\omega\varepsilon_{\alpha\gamma\delta}A_{\gamma\delta\beta}$ . The first term in equation 2.22 is the sum of the pairs of bonds that make up the chiral structures (from the two-group mechanism), the second term is due to the changes in the position vector  $R_j$  for a group with respect to the molecular origin (from the inertial mechanism) and the last term is a product of the intrinsic bond polarizability and optical activity tensors.

Assuming that the bonds are achiral and have axial symmetry, the last term in the above expression is equal to zero since the intrinsic optical activity tensors  $G'_{i\alpha\beta}$  and  $A_{i\alpha\beta\gamma}$  are zero (Barron, 1982; Barron & Clark, 1982). The isotropic contribution  $\langle 0 | \alpha_{\alpha\alpha} | 1_p \rangle \langle 1_p | G'_{\beta\beta} | 0 \rangle$  can also be assumed to be zero. This leads to the conclusion that in this situation of axially symmetric achiral bonds, the magnetic dipole and electric quadrupole contributions are equal. This leads to some convenient simplifications that will now be discussed.

## 2.6 Backscattered ROA

Utilising the results from the general bond polarizability theory that  $\alpha G' = 0$  and  $\beta (G')^2 = \beta (A)^2$  in equations 2.10 it is found that (Barron, 1982; Hecht *et al.*, 1989)

$$\Delta(0^\circ) = 0 \quad (2.23a)$$

$$\Delta(180^\circ) = \frac{64\beta (G')^2}{2c[45\alpha^2 + 7\beta(\alpha)^2]} \quad (2.23b)$$

$$\Delta_x(90^\circ) = \frac{16\beta (G')^2}{c[45\alpha^2 + 7\beta(\alpha)^2]} \quad (2.23c)$$

$$\Delta_z(90^\circ) = \frac{8\beta (G')^2}{6c\beta(\alpha)^2} \quad (2.23d)$$

Therefore the ROA intensity in the forward direction is zero and it is maximised in the backward direction (where it is four times that for polarized right-angle scattering).

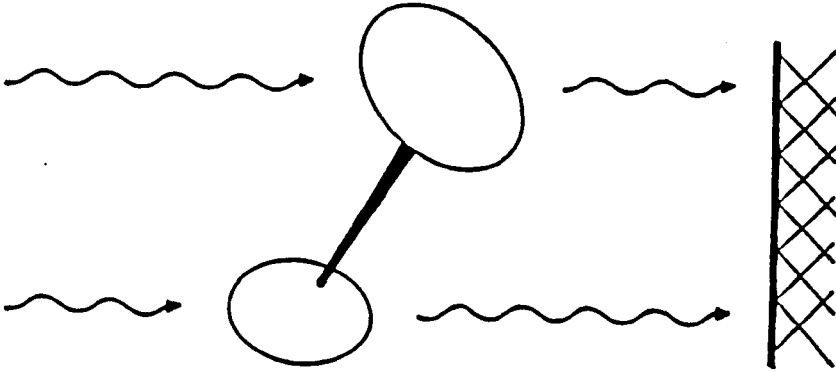


Figure 2.3 There is no net ROA in the forward direction in the two-group model because there is no phase difference between the two scattered waves.



This result is illustrated in the two-group model shown in Figure 2.3 where it can be deduced that there will be no ROA intensity in the forward direction because both waves interfering at the detector will have the same optical path distance. However it is not immediately apparent from these simple pictures that ROA is maximised in the backscattering direction. These results led to the conclusion that backscattering is the optimum experimental arrangement which must be used for studying unfavourable samples such as biopolymers in aqueous solution.

## 2.7 Ab Initio Calculations

*Ab initio* calculations have recently been used to predict ROA intensities (Bose *et al.*, 1990; Polavarapu, 1990; Polavarapu, 1993a,b). Although no molecules larger than alanyl alanine have yet been treated by this method it is worth giving a summary of the method used because the calculations are becoming more important.

The method is based on the previously discussed Placzek polarizability approximation. It therefore requires the computation of polarizability and optical activity tensor derivatives such as  $(\delta\alpha_{\alpha\beta}/\delta Q_p)_0$ ,  $(\delta G'_{\alpha\beta}/\delta Q_p)_0$  and  $(\delta A_{\alpha\beta\gamma}/\delta Q_p)_0$  to calculate the ROA intensities. Since the scattering is at transparent frequencies a reasonable assumption of replacing  $\omega_{jn}^2 - \omega^2$  by  $\omega^2$  is made so that the polarizability and optical activity tensors now simplify to

$$\alpha_{\alpha\beta} = 2 \sum_{j \neq n} \frac{1}{W_{jn}} \text{Re}(\langle n | \mu_\alpha | j \rangle \langle j | \mu_\beta | n \rangle) \quad (2.24a)$$

$$G'_{\alpha\beta} = -\frac{2\omega}{\hbar} \sum_{j \neq n} \frac{1}{W_{jn}^2} \text{Im}(\langle n | \mu_\alpha | j \rangle \langle j | m_\beta | n \rangle) \quad (2.24b)$$

$$A_{\alpha\beta\gamma} = 2 \sum_{j \neq n} \frac{1}{W_{jn}} \text{Re}(\langle n | \mu_\alpha | j \rangle \langle j | \Theta_{\beta\gamma} | n \rangle) \quad (2.24c)$$

$W_{jn}$  represents the energy difference between the  $j$ th virtual intermediate state ( $W_j$ ) and the initial state ( $W_n$ ).

The polarizability is dealt with initially. This first of all involves considering the wavefunction in a static electric field (Amos, 1982)

$$\psi_n(E_\beta) = \psi_n^{(0)} + E_\beta \psi_n^{(1)}(E_\beta) + \dots \quad (2.25)$$

where  $\psi_n^{(1)}$  from perturbation theory is

$$\psi_n^{(1)}(E_\beta) = \sum_{j \neq n} \frac{1}{W_{jn}} \langle j | \mu_\beta | n \rangle | j \rangle \quad (2.26)$$

The polarizability can then be expressed as

$$\alpha_{\alpha\beta} = 2 \langle \psi_n^{(0)} | \mu_\alpha | \psi_n^{(1)}(E_\beta) \rangle \quad (2.27)$$

By writing the wavefunctions in terms of molecular orbitals  $\phi_k$  perturbed by the static electric field, the final computational version is

$$\alpha_{\alpha\beta} = 4 \sum_{k,occ} \langle \phi_k^{(0)} | \mu_\alpha | \phi_k^{(1)}(E_\beta) \rangle \quad (2.28)$$

Applying the same technique to the electric dipole-electric quadrupole optical activity tensor we find

$$A_{\alpha\beta\gamma} = 4 \sum_{k,occ} \langle \phi_k^{(0)} | \Theta_{\beta\gamma} | \phi_k^{(1)}(E_\alpha) \rangle \quad (2.29)$$

Due to the fact that the electric dipole-magnetic dipole optical activity tensor  $G'_{\alpha\beta}$  vanishes as  $\omega \rightarrow 0$  and therefore does not have a static limit it has to be treated in a slightly different manner. Since  $(1/\omega) G'_{\alpha\beta}$  does have a static limit (Amos, 1982) and

if  $\psi_n^{(1)}(\beta_\beta)$  is the wavefunction perturbed in a static magnetic field we can write

$$[(1/\omega) G'_{\alpha\beta}]_{\omega=0} = -2\hbar \text{Im}(\langle \psi_n^{(1)}(E_\alpha) | \psi_n^{(1)}(B_\beta) \rangle) \quad (2.30)$$

In terms of the perturbed molecular orbitals

$$[(1/\omega) G'_{\alpha\beta}]_{\omega=0} = -4\hbar \sum_{k, \text{occ}} \text{Im}(\langle \phi_k^{(1)}(E_\alpha) | \phi_k^{(1)}(B_\beta) \rangle) \quad (2.31)$$

It is now possible to obtain the polarizability and optical activity tensors from two coupled Hartree-Fock SCF calculations, one giving the molecular orbitals perturbed by a static electric field and the other by a static magnetic field.

The required derivatives such as  $(\delta\alpha/\delta Q_p)_0$  are then calculated numerically by using the above results to evaluate  $\alpha_{\alpha\beta}$ ,  $(1/\omega) G'_{\alpha\beta}$  and  $A_{\alpha\beta\gamma}$  in their static limits at the equilibrium geometry and at the geometries displaced by 0.005 Å along each atomic coordinate. These calculations are performed by the CADPAC program (Amos & Rice, 1987). This method is restricted at the moment to less than 80 electrons.

## Chapter 3

### Instrumentation

In this chapter the design of the new backscattering ICP Glasgow University ROA spectrometer No.3 (GUROAS3) will be discussed.

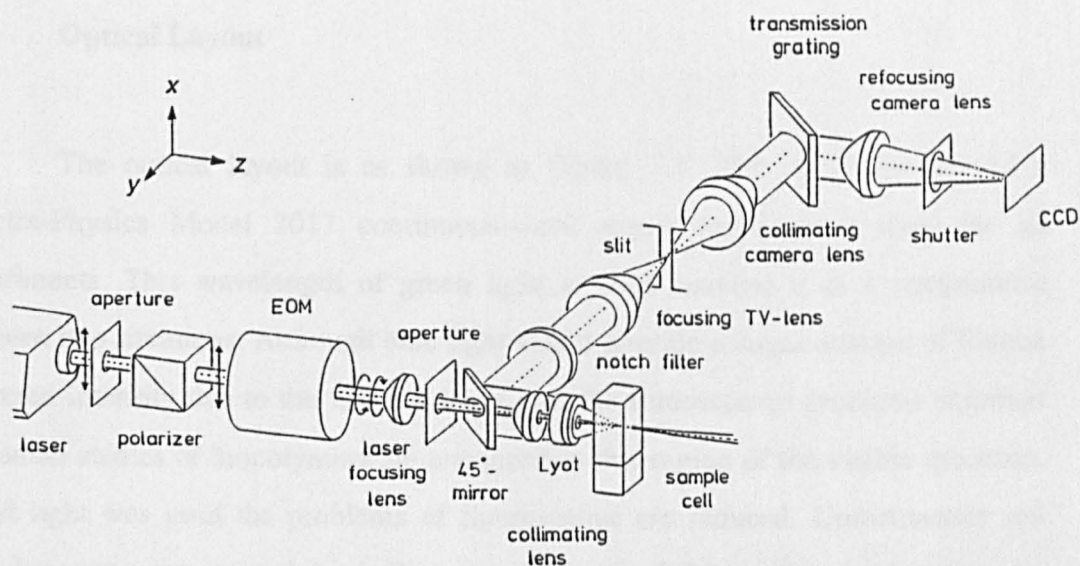


Figure 3.1 The optical layout of the backscattering GUROAS3

The instrument is based on a backscattering geometry since as discussed earlier, backscattered ROA data exhibit a much larger signal-to-noise ratio as compared to

right-angle scattering (Barron *et al.*, 1989, 1990; Hecht *et al.*, 1989, 1992; Hecht & Barron, 1990, 1994; Hecht & Nafie, 1991; Barron & Hecht, 1993).

Each component of the instrument will now be described in detail. The major improvement compared to GUROAS2 was the introduction of a low F/number stigmatic (imaging) spectrograph (Battey *et al.*, 1993; Purcell, 1993; Tedesco *et al.*, 1993) as compared to the previously used astigmatic (non-imaging) spectrographs that were used in GUROAS2 and GUROAS1. It should be pointed out at this stage that the construction of the instrument was the work of L. D. Barron and L. Hecht.

### 3.1 Optical Layout

The optical layout is as shown in Figure 3.1. The 514.5 nm line of a Spectra-Physics Model 2017 continuous-wave argon ion laser is used for all experiments. This wavelength of green light is used because it is a compromise between two situations. Although blue light would provide a larger amount of Raman scattered intensity due to the  $1/\lambda^4$  Rayleigh law, the fluorescence problems common in Raman studies of biopolymers are enhanced in this region of the visible spectrum. If red light was used the problems of fluorescence are reduced. Unfortunately red light does not create enough basic Raman intensity for ROA studies at this time to be feasible. The use of green light is thus a compromise between fluorescence suppression and the  $1/\lambda^4$  Rayleigh law.

The green laser beam, of diameter  $\sim 2$  mm, first passes through an aperture which helps to suppress plasma background radiation. To increase the degree of polarization the light then passes through a polarizer. A Glan Taylor laser prism polarizer is used for this purpose (Klinger *et al.*, 1990) which has an extinction ratio of  $10^{-5}$  in the visible region. The prism has anti-reflection coatings on both its faces and provides about a 0.95 transmittance for green light.

The linearly polarized laser light then passes into the electro-optic modulator (EOM). By applying a high voltage across a potassium dideuterium

phosphate crystal and reversing the direction of the electric field near-perfect right and left circular polarization can be achieved. This is an improved model from that used in GUROAS1 and GUROAS2, and is based on a special electrode combination instead of the usual ring electrodes. This creates a more uniform electric-field in the modulator crystal and therefore will minimise any stress-induced birefringence. The EOM is temperature controlled at  $26.0 \pm 0.1^\circ\text{C}$ , as the quarter-wave voltage is a function of temperature. A Leysop high-voltage linear differential amplifier is used to drive the EOM, creating a highly stable bipolar high-voltage square wave. This then allows the EOM to switch phase and generate right and left circularly polarized light. The amplifier also has adjustable positive and negative voltage limits that are sometimes required to compensate for the differences in birefringence between different samples. The EOM is connected to a computer *via* a parallel interface so that the modulation frequency can be altered. There is also an electronics module that not only starts each modulation in the right phase but also resets the voltage along the crystal to zero after each cycle. The actual alignment of the EOM is one of the most critical factors in the ROA instrument; if it is incorrectly aligned the ROA spectrum will contain artifacts. Strongly polarized Raman bands tend to be most susceptible to artifacts. A Soleil-Babinet compensator is used to allow the quarter-wave voltages for the EOM to be set up as precisely as possible.

The highly pure circularly polarized laser beam is then focused *via* a plano-convex synthetic fused-silica lens. The beam of light then passes through another aperture to cut out any stray reflected light.

The light then passes through a  $45^\circ$  mirror, a collimating lens and a Lyot depolarizer. These three components all contain a hole in the middle and the beam passes straight through until it hits the sample cell (Hecht *et al.*, 1992).

The sample cell is a quartz (Spectrosil) microfluorescence cell (Optiglass) that measures  $5 \times 6 \times 12.5$  mm. The cell has a maximum capacity of  $\sim 300 \mu\text{l}$  but in practice  $\sim 125 \mu\text{l}$  of a sample is usually used. The sample cell is also annealed at  $\sim 1200^\circ\text{C}$ , to remove stray birefringence in the glass which can cause artifacts. The sample cell is held in an optical mount (Newport) that keeps the cell fixed throughout

the duration of the experiment. The optical mount allows the cell to be moved precisely in any of the x, y or z-directions. The actual alignment of the cell is critical, since the back-reflections from the cell walls should be reflected back along the central axis of the optics. If this is not performed properly and the light is reflected into the collection optics of the spectrograph, artifacts and problems with the baseline of the spectra will arise.

The backscattered cone of Raman light from the sample then passes through a Leysop calcite Lyot depolarizer (Mochizuki, 1984; Takada *et al.*, 1988; Kliger *et al.*, 1990) to scramble any residual linearly polarized components. The Lyot is made of two calcite plates of different thickness, that are cut from the same X-cut of a raw calcite crystal. These two plates are then glued together using optical cement with their optical axes oriented at  $45^\circ$  to each other. Both faces of the Lyot are anti-reflection coated. The Lyot can be rotated around the z-direction so that the optimum depolarization position can be found *via* a trial and error procedure.

The backscattered light is then collimated with a symmetric convex synthetic fused-silica lens, which again contains anti-reflection coatings. The light is then deflected into the  $90^\circ$  (-y) direction by a  $45^\circ$  oriented mirror that has an enhanced silver coating which was made by Balzers. This mirror is a considerable improvement on the previously used aluminium mirrors. The silver mirror has a difference in reflection coefficients of horizontal and vertically polarized light of 0.5 - 0.6% for visible light. For aluminium coated mirrors the value is 5 - 6% and thus represents a ten-fold improvement in suppression of this particular source of artifacts (Hecht *et al.*, 1992; Bell *et al.*, 1993).

The deflected light at this point contains a mixture of Raman and Rayleigh light. The Rayleigh light if not suppressed will have a detrimental effect on the signal-to-noise ratio and will introduce slopes into the ROA baseline. To reject the stray Rayleigh light a holographic notch filter is used. A major improvement over GUROAS1 where the notch filter was placed in front of the entrance slit of the spectrograph is that the notch filter in the new instrument can operate in quasi-parallel light (Yang *et al.*, 1991) and thus has a much greater efficiency and rejection

of stray light. A Super Notch-Plus filter with an optical density  $> 8$  is currently used which allows the collection of Stokes and Anti-Stokes data down to  $50 \text{ cm}^{-1}$ . Figure 3.2 is a representation of how the Super Notch-Plus filter works by rejecting the Rayleigh light at  $515.5 \text{ nm}$ . The transmission of the filter can be observed to be over 80% in the wavelength spread corresponding to our Raman spectra. It must also be noted that the Kaiser and older versions of the spectrographs are all single grating spectrographs, which is only made possible due to the excellent stray light rejection characteristics of present day holographic notch and edge filters.

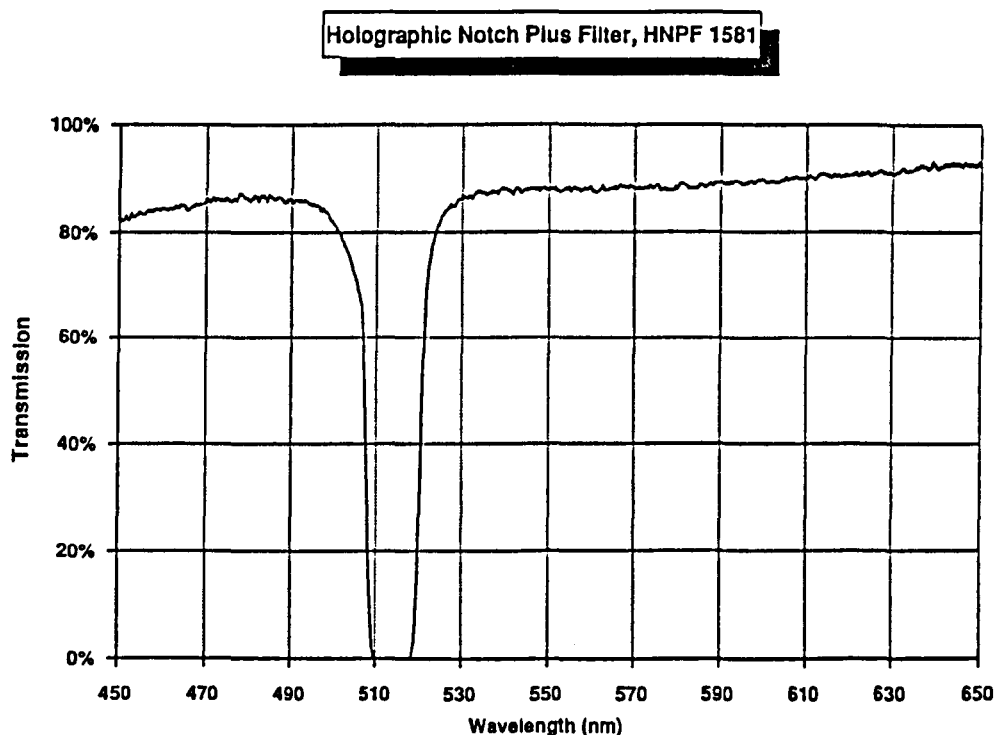


Figure 3.2 The performance of the holographic notch filter



After passing through the holographic notch filter the Raman light is focussed by a F/1.4 CC TV-lens onto the vertical entrance slit of the spectrograph. Probably the greatest improvement in the instrument has been the use of a new spectrograph which will now be discussed.

The Holospec F/1.8i from Kaiser Optical Systems is based on the latest stigmatic spectrograph technology. The first half of the commercial version uses a preconfigured filter stage which would have involved the use of extra and unnecessary optics. It was therefore decided to use only the second half of the spectrograph in GUROAS3. An advantage of the new spectrograph is that an acceptance cone of F/1.4 is maintained throughout the spectrometer, eliminating the requirement of any further refocusing lenses. It is probably not an exaggeration to say that, due to this stigmatic spectrograph, GUROAS3 is the worlds fastest Raman spectrograph. The major improvement in the design of the new spectrograph is that it operates *via* a novel holographic transmission grating as opposed to a conventional ion-etched reflection grating as used in GUROAS1 and GUROAS2. The transmission grating placed between the collimating and refocusing camera lens is based on new volume-technology (Tedesco *et al.*, 1993) and produces a 90° fold of the optical path and has an efficiency of greater than 80% for unpolarized light (Battey *et al.*, 1993; Purcell, 1993). This is a great improvement over the previously used ion-etched gratings which only have an efficiency of 60%. The transmission grating has 2,400 grooves mm<sup>-1</sup> which results in a reciprocal linear dispersion of 3.0 nm mm<sup>-1</sup>, providing a spectral coverage of about 1,200 cm<sup>-1</sup> across the CCD chip in the detector.

The use of a transmission grating is an integral part of the design since this allows an on-axis spectrometer set-up to be used, providing nearly diffraction-limited imaging (stigmatic) performance. In GUROAS1 and GUROAS2 an off-axis concave camera and collimating mirrors in the old spectrograph (operating *via* an in-plane asymmetric Czerny-Turner optical configuration) produced unwanted astigmatism. The end result is that an approximate fivefold increase in throughput is obtained as compared to the old astigmatic spectrograph.

As a detector in the instrument a backthinned charge-coupled device (CCD) is employed (Billhorn *et al.*, 1987a,b) which is now becoming common for Raman spectroscopy (Dierker *et al.*, 1987; Pemberton & Sobocinsky, 1989; Williamson *et al.*, 1989; Knobloch *et al.*, 1989). A Wright Instruments CCD is used which contains a thermoelectrically cooled camera head, a camera electronics module and a computer board. The camera head contains a single-field area array CCD image sensor which comprises a two-dimensional array, measuring 385 columns by 578 rows of square pixels. Each of these pixels can store a photo-induced electronic charge. The CCD is cooled to 200K using a Peltier effect cooler, allowing the detector to have an excellent signal-to-noise ratio with a dark current (thermally excited signals) of only about 0.14 electron counts per pixel per second. The CCD chip is thinned by an acid etching process to improve its efficiency. The CCD is oriented in such a way as to maximise the spectral coverage of the Raman spectrum; and is therefore oriented with its short axis (385 pixels) along the slit axis of the spectrograph and the long axis (578 pixels) parallel to the direction of dispersion of the spectrometer.

The collection of data on the CCD chip works using a serial pixel binning (Epperson *et al.*, 1990) process which basically converts the two-dimensional CCD image detector into a linear array.

The end result is that the backthinned CCD detector provides a light detection unit with a quantum efficiency in excess of 80% in the 450 to 650 nm range. Standard CCD's have a quantum efficiency of about 20% at 500 nm and about 50% at 700 nm. CCD's also have extremely low noise levels and so, unlike diode arrays, they can be used without intensification. An illustration of the efficiency of CCD's is shown in Figure 3.3.

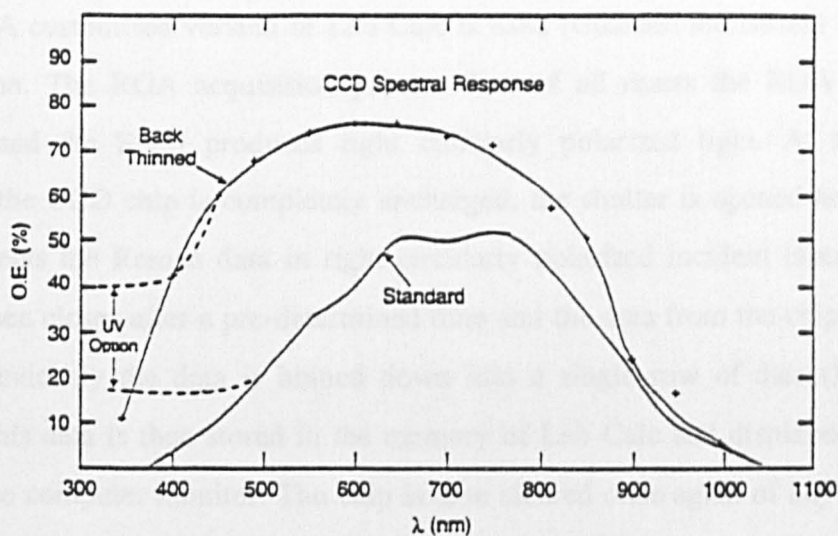


Figure 3.3 Efficiency of the backthinned CCD detector

### 3.2 Computer Control

A Dell personal computer uses a computer interface to synchronise the CCD detector and the EOM. Figure 3.4 shows the computer and electronic set-up of the ROA instrument.

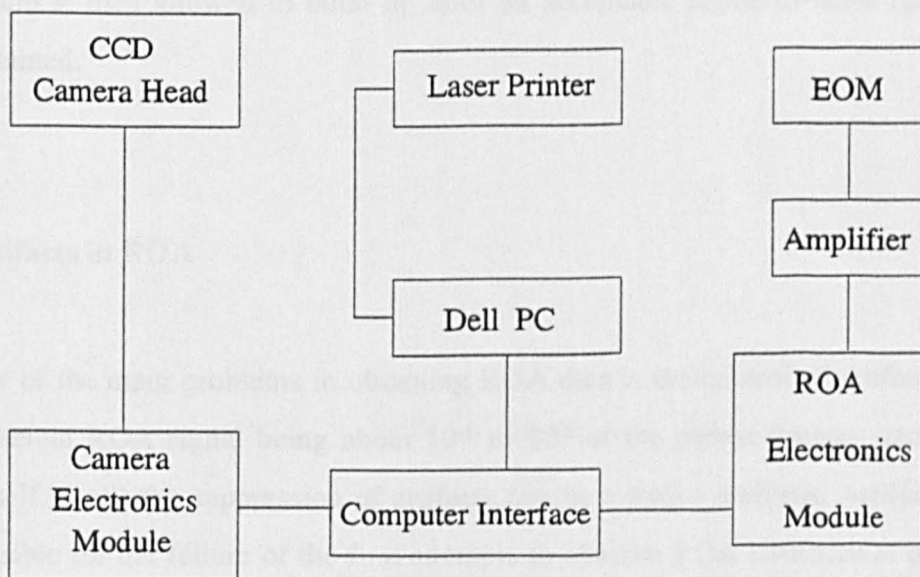


Figure 3.4 Computer control of the ROA instrument

A customised version of Lab Calc is used (Galactic Industries) for the data acquisition. The ROA acquisition process first of all resets the ROA electronics module and the EOM produces right circularly polarized light. At this precise moment the CCD chip is completely uncharged, the shutter is opened and the CCD chip collects the Raman data in right circularly polarized incident laser light. The shutter then closes after a pre-determined time and the data from the chip is read. As stated previously the data is binned down into a single row of data (Lacy *et al.*, 1991). This data is then stored in the memory of Lab Calc and displayed in the top half of the computer monitor. The chip is then cleared once again of any charge and the computer commands the EOM to produce left circularly polarized light. The shutter is opened once more and the CCD chip collects the Raman data in left circularly polarized incident light for exactly the same time of collection as in right. This left spectrum is stored in the Lab Calc memory and displayed in the top half of the monitor, in a different colour to that of the right. The left Raman spectrum is then subtracted from the right Raman spectrum and the resultant ROA spectrum displayed in the bottom half of the monitor. The memory containing the right and left Raman spectrum is then deleted and the modulation process starts again. After each cycle the obtained difference spectrum is added to the previously obtained ROA spectrum. The ROA spectrum is then allowed to build up until an acceptable signal-to-noise ratio has been obtained.

### 3.3 Artifacts in ROA

One of the main problems in obtaining ROA data is the control of artifacts. Due to the actual ROA signal being about  $10^{-3}$  to  $10^{-5}$  of the parent Raman signal (which is itself small) the suppression of artifacts can be a major problem. Artifacts were responsible for the failure of the first attempts to observe ROA (Bosnich *et al.*, 1972; Diem *et al.*, 1973) and remain a problem even to this day.

It was only after the source of artifacts was understood were the first ROA

data collected (Barron *et al.*, 1973). The source of artifacts can be understood from the polarization dependence of the isotropic and anisotropic molecular polarizability contributions to the scattering intensity at  $90^\circ$ . It was found that strongly polarized Raman bands tend to be much more susceptible to artifacts than depolarized bands. Artifacts themselves are 'ghost' signals that can occur as 'S' shaped signals on strongly polarized Raman bands and have absolutely nothing to do with the conformation or chirality of the molecule being examined (there are also bias artifacts which occur when the spectra is not properly on the baseline). The actual analysis of artifacts has now become quite sophisticated with the use of phenomenological Stokes Mueller formalisms (Hecht & Jordanov, 1987; Kliger *et al.*, 1990).

Utilising this information from the Stokes Mueller theoretical formalisms, residual linear birefringence in the optical components was deduced to be a major factor leading to the generation of artifacts. To reduce this stray birefringence high quality strain-free quartz is used wherever possible in the instrument.

As stated previously the alignment of the EOM is crucial to the control of artifacts; a Soleil-Babinet compensator is used for the correct alignment of this component. The EOM crystal is also temperature stabilized and supplied with a highly stable voltage by a linear differential amplifier, in an attempt to suppress artifacts.

Backscattering has the advantage over right-angle scattering that artifacts are greatly reduced due to the backscattered light having a near-perfect rotational symmetry (Escribano, 1985; Hecht & Barron, 1990). This has the effect of cancelling out most of the artifacts that are dependant on the azimuthal scattering angle (Hecht & Barron, 1990).

Another source of artifacts is that back-reflected light from elements in the optical train and the sample cell passing into the laser cavity can degrade the degree of polarization of the incident laser light. To limit this problem anti-reflection coatings are used wherever possible. Other less important sources of artifacts include optical rotation of linearly polarized contaminants, which can be limited by using

smaller sample cells. Dust particles can also be a problem and these are eliminated by filtering and centrifuging the samples for at least 10 minutes.

The Lyot depolarizer is an essential part of the instrument because if it was not used the ROA spectrum would still have many artifacts even if all of the above precautions were taken. The Lyot depolarizer is essential because the  $45^\circ$  mirror introduces huge polarization artifacts if the Raman light is not effectively depolarized (due to residual linear contaminants). The Lyot depolarizer has different depolarization characteristics at different orientations and the correct orientation is found by trial and error. The dipeptide L-alanyl-L-alanine is used as a standard to find the correct orientation of the Lyot. The dipeptide contains a strongly polarized band at  $886\text{ cm}^{-1}$  and when the Lyot is in the approximately correct position, small changes in orientation can eliminate or create a large S-shaped signal on this band (Barron *et al.*, 1990; Barron *et al.*, 1992a). It is found that for proteins and polypeptides the optimum orientation of the Lyot can differ by  $10 - 20^\circ$  from the position for peptides. A high quality spectrum of L-alanyl-L-alanine is shown in Figure 3.5, where the S-shaped curve on the  $886\text{ cm}^{-1}$  band has been suppressed in this way.

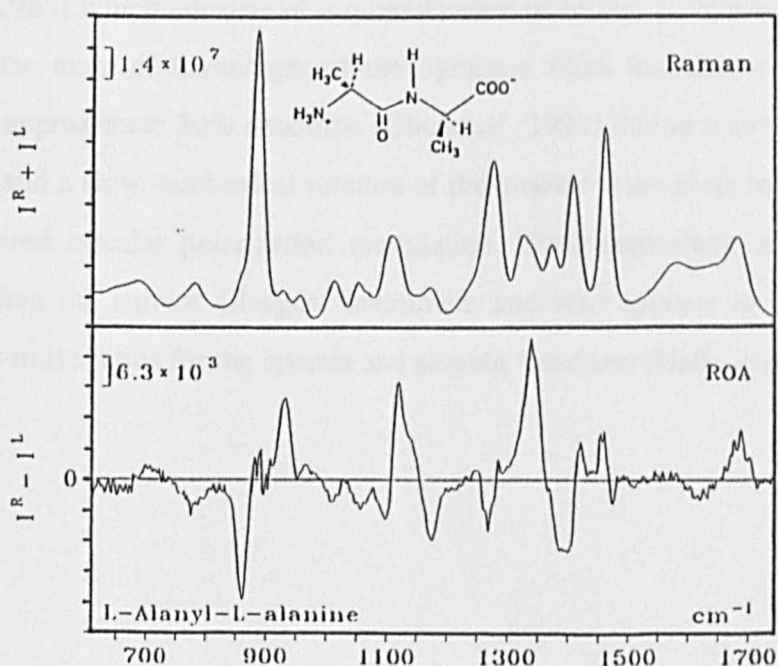


Figure 3.5 Backscattered Raman and ROA of L-alanyl-L-alanine

### 3.4 Instrument Performance

The new ROA instrument is approximately five times faster for small molecules such as pinene but only about two to three times faster for a protein due to the protein having a more disperse scattering geometry. In Figure 3.6 the backscattered ROA spectra of the two enantiomers of trans-pinane are shown. The ROA spectra of both enantiomers are displayed after 10 minutes of collection. In the bottom part of Figure 3.6 the ROA spectra of (+)-trans-pinane is shown after 20 seconds of collection. Even after this short time of collection all of the major ROA signals can be observed (it should be noted that these spectra were collected on the old spectrograph).

The remarkable speed of the instrument can be appreciated when it is compared with other ROA instruments around the world. In Syracuse a backscattering ROA instrument using an unthinned CCD detector was built (Nafie *et al.*, 1991; Che *et al.*, 1991; Che & Nafie, 1992) that was capable of measuring the ROA spectra of amino acids, peptides and saccharides (Nafie *et al.*, 1991) using a DCP modulation technique based on a conventional circular analyser (Bickel & Bailey, 1985) which consists of a quarter-wave plate and a dichroic sheet polarizer. One of the main disadvantages of the Syracuse ROA instrument is that it suffers from an approximate 50% deadtime (Che *et al.*, 1991) due to a slow read-out of the detector and a slow mechanical rotation of the quarter-wave plate required to achieve the required circular polarization modulation. Their instrument is therefore much slower than the current Glasgow instrument and also appears to suffer from poor signal-to-noise ratios for the spectra and sloping baselines (Nafie *et al.*, 1991).



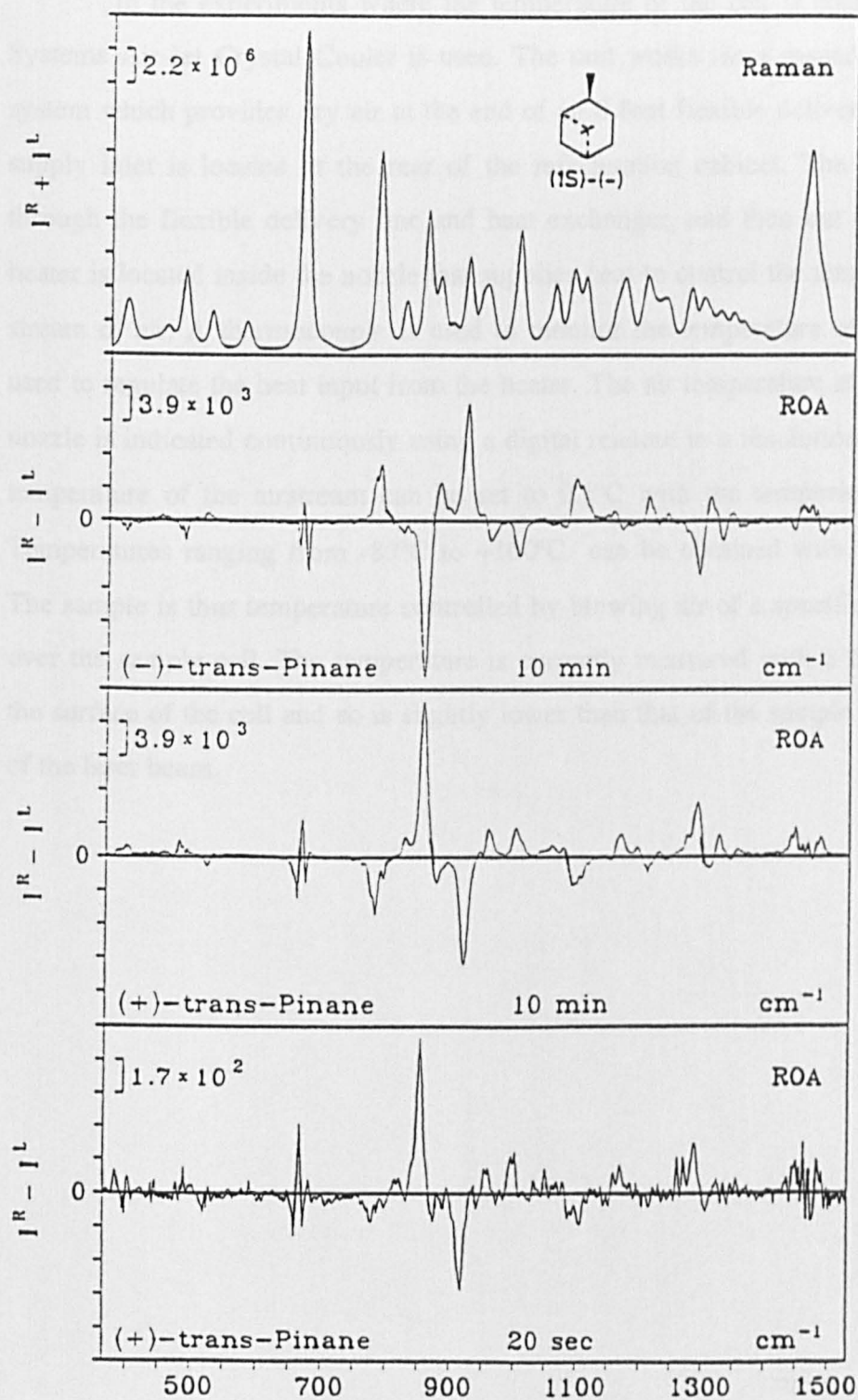


Figure 3.6 Backscattered ROA spectra of (+) and (-) -trans-pinane



### 3.5 Temperature Control of Sample Cell

In the experiments where the temperature of the cell is controlled an FTS Systems Air-Jet Crystal Cooler is used. The unit works *via* a cascade refrigeration system which provides dry air at the end of an 8 foot flexible delivery line. The air supply inlet is located at the rear of the refrigeration cabinet. The air then flows through the flexible delivery line and heat exchanger, and then out of a nozzle. A heater is located inside the nozzle that supplies heat to control the temperature of the stream of air. A thermocouple is used to monitor the temperature of the air and is used to regulate the heat input from the heater. The air temperature at the end of the nozzle is indicated continuously using a digital readout to a resolution of 0.1°C. The temperature of the airstream can be set to 0.1°C with the temperature controller. Temperatures ranging from -85°C to +100°C can be obtained with this apparatus. The sample is thus temperature controlled by blowing air of a specified temperature over the sample cell. The temperature is currently measured with a thermometer at the surface of the cell and so is slightly lower than that of the sample along the axis of the laser beam.

## Chapter 4

### Protein Structure Determination

In this chapter a brief outline will be provided of the different techniques that are currently employed for protein structure determination, with an indication of where and how ROA can provide new information.

#### 4.1 Protein Structure

A protein consists of many amino acids and the conformation of these subunits are defined *via* the Ramachandran phi ( $\phi$ ) and psi ( $\psi$ ) angles (Figure 4.1). The overall conformation of the protein will be determined by the combination of all  $\phi$  and  $\psi$  angles.

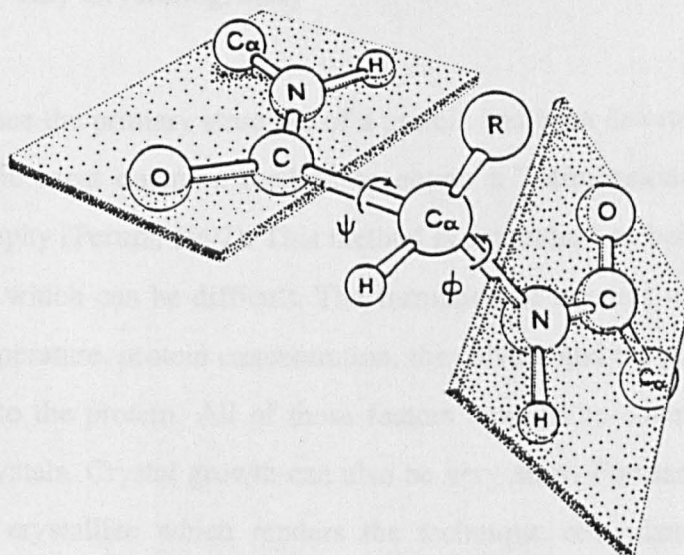


Figure 4.1 Illustration of the peptide backbone

There are four levels of protein architecture that will now briefly be explained. The primary structure is the amino acid sequence and a list of any disulfide bridges between two cysteines. It is thus a complete map of the covalent bonded network of the protein. The secondary structure is a description of the spatial arrangement of adjacent amino acids. If regular arrangements extend over a number of residues this can lead to  $\alpha$ -helix ( $\phi \approx -60^\circ$ ,  $\psi \approx -50^\circ$ ) and  $\beta$ -sheet ( $\phi \approx -120^\circ$ ,  $\psi \approx +130^\circ$ ) structures developing. Supersecondary structure refers to clusters of secondary structure. An example of supersecondary structure is where a  $\beta$ -strand is separated from another  $\beta$ -strand by an  $\alpha$ -helix; this is called a  $\beta$ - $\alpha$ - $\beta$  motif (Branden & Tooze, 1991). Tertiary structure is the compact globular units formed by the packing of  $\alpha$ -helices and  $\beta$ -sheets associated with loops and turns. If there is more than one polypeptide chain (i.e. subunit) in the protein the quaternary structure refers to the spatial arrangement of these subunits and the nature of their contacts.

## 4.2 X - Ray Crystallography

Once the primary structure of a protein has been determined by biochemical methods, the most common method to obtain a 3-dimensional structure is X-ray crystallography (Perutz, 1992). This method is dependent on being able to crystallize the protein which can be difficult. The formation of crystals is critically dependent on pH, temperature, protein concentration, the solvent and the presence of added ions or ligands to the protein. All of these factors have to be controlled to produce the required crystals. Crystal growth can also be very slow. Furthermore, many proteins will never crystallize which renders the technique redundant for these proteins. Proteins seem to be difficult to crystallize if they have large mobile regions, such as unstructured polypeptide tails or carbohydrate moieties.

There is also the debate that crystal and solution structures may not be the same. Crystal contacts may perturb protein structures and sometimes provide an

incorrect description of a protein surface. The chance to perturb the conformation globally by crystal contacts is more significant in small rather than in large proteins. Nonetheless even in large proteins the active sites of the proteins often reside on protein surfaces that might therefore be perturbed by crystallization.

Even when crystals are obtained they may not diffract properly and may provide an intractable diffraction pattern. However, despite these problems X-ray crystallography remains the best method for protein structure determination.

## 4.2 Multi - Dimensional Nuclear Magnetic Resonance (NMR)

In recent years NMR has evolved into a powerful technique for protein structure determination. As opposed to the previously described X-ray method, these experiments are conducted in solution and therefore provide a superior insight into the actual structure of the protein under near physiological conditions.

There are however major problems associated with multi-dimensional NMR and protein structure determination. The most difficult one to overcome is the 'molecular weight barrier'. The largest structures actually solved by NMR are only about 18 kDa in size (Clare *et al.*, 1991; Fairbrother *et al.*, 1992), although resonance assignments for proteins as large as 22 kDa for monomeric proteins (Stockman *et al.*, 1992) and even 31 kDa for dimeric proteins (Grzesiek *et al.*, 1992) have been reported. It therefore gets too difficult to interpret the spectra of large molecules due to the very large number of nuclear overhauser enhancement (NOE) effects involved.

Solving a protein structure by NMR can also depend on aspects other than molecular size. The main factor is whether a protein exhibits good NMR spectra. It is unfortunate that only a small number of the proteins that fall within the suitable molecular weight range provide an NMR spectra that can be assigned. Even if assignments can be made, there is no guarantee that a precise structure can be obtained. It is also found that some proteins just show fewer NOEs than others, independent of their size.

The NMR spectra tend to be very crowded due to the very large number of resonances. This peak overlap problem has been solved to a large extent by the development of multidimensional and multiple resonance experiments, at least for proteins up to 25 kDa that can be isotope labelled. Aromatic resonances still suffer from these problems, making these regions notoriously difficult to assign.

The fast transverse relaxation rates of large proteins also impose limits to the lengths of the pulse sequences that can be applied. This leads to line broadening in the protein NMR spectra. Increasing the size of the protein being investigated also limits the maximum concentration that can be used and thus restricts the sensitivity of the technique. At high concentrations the onset of protein aggregation adds to the line broadening problem. The electrostatic or hydrophobic interactions that cause this effect can severely broaden the spectrum to such an extent as to make the obtained spectra of little use.

It has also been found that large proteins often have significant moieties that do not show NOEs and it is not uncommon for some residues not to have any detectable resonances. This problem is usually due to multiple conformations in intermediate exchange. Due to the slow timescale involved ( $\sim 10^{-3}$  s) this is a serious problem in NMR but relaxation time experiments can provide valuable tests to see if mobility is causing problems for structure analysis. To solve the structure of large proteins isotope labelling with  $^{15}\text{N}$  and  $^{13}\text{C}$  is an essential procedure. Heteronuclear multidimensional NMR experiments can then be used for the structure determination.

#### 4.3 Computer Simulations of Protein Structures

Recent advances in computer programs are providing a third way to calculate the structure of a protein from the amino acid sequence (Chan & Dill, 1993; Borman, 1995). This is a purely theoretical approach and is currently showing significant promise.

One of these computer programs can predict secondary structure (i.e.  $\alpha$ -

helices and  $\beta$ -strands) and supersecondary structure (i.e. linked helices and strands) but has not yet been successful in predicting the tertiary structure (i.e. the overall structure of the folded protein). The other type of program (Dill & Chan, 1993) can provide a prediction of the tertiary structure but not the secondary structure. There remains much scope for great advancement in this area though the limiting factor at the moment is the very large number of parameters involved in the calculations.

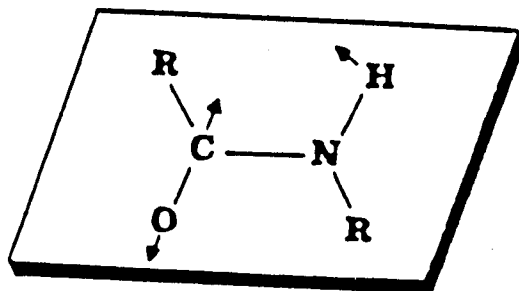
#### 4.4 Application of Spectroscopic Techniques to Biopolymers

Currently ECD is the most commonly used spectroscopic technique to characterise biopolymers and has been applied in both peptide and protein structure studies (Woody, 1977). However, ECD structure analyses are limited by the number and resolution of transitions that are available. The most useful electronic transitions that have been found are the  $\pi^* \leftarrow n$  and  $\pi^* \leftarrow \pi$  transitions of the amide groups. These provide information about the polymeric backbone but, since such amide electronic excitations are relatively delocalised, the resulting UV bands are often affected by other environmental or local perturbations. ECD is also limited by its lack of sensitivity to certain types of structure, such as the tertiary structure of proteins. Nonetheless, far-UV CD can be used to detect both  $\alpha$ -helix and  $\beta$ -sheet conformations. Analysis of these bands can also reveal the actual quantity of each type of structure present. Near-UV CD can also be used to monitor the environments of aromatic residues, such as tryptophans.

Vibrational techniques have also been applied extensively to biopolymers. Vibrational spectroscopy involves several distinct vibrational modes involving the -CO-NH- group, the most important of which will now be explained (Tu, 1986). Vibrations of the peptide backbone usually occur in four main regions of the Raman spectrum (Tu, 1986): the  $C_\alpha$ -C stretch region  $\sim 870 - 960 \text{ cm}^{-1}$ , the  $C_\alpha$ -N stretch region  $\sim 1020 - 1150 \text{ cm}^{-1}$  the amide III region  $\sim 1230 - 1310 \text{ cm}^{-1}$  and the amide I region  $\sim 1630 - 1700 \text{ cm}^{-1}$ . The amide III region was thought to be mainly due to N-H

in-plane deformations coupled with  $C_\alpha$ -N stretch modes but in fact Diem *et al.* (1992) showed by examining small peptides such as alanyl alanine and related isotopomers that the amide III region actually involves much more mixing between the N-H and  $C_\alpha$ -H modes than was previously thought. This conclusion is supported by a recent ROA study of alanylpeptide oligomers (Ford *et al.*, 1994).

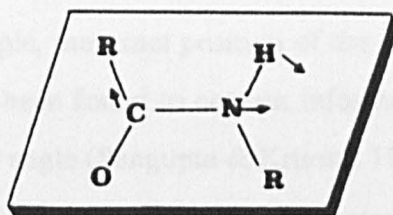
The amide I vibrational mode has its main contribution from the in-plane peptide  $C=O$  stretching vibration (80%) and a smaller contribution from the N-H in-plane bending vibration (Figure 4.2 ).



Amide I

Figure 4.2 The amide I vibrational mode

The amide III band (Figure 4.3) on the other hand does not involve the  $C=O$  vibration.



Amide III

Figure 4.3 The amide III vibrational mode

The amide II vibration ( $\sim 1510 - 1570 \text{ cm}^{-1}$ ) is dominated by the N-H in-plane deformation, with a much smaller contribution from the C-N stretching mode (Bandeekar, 1991).

The amide I and III bands have relatively high Raman intensity and correlations with protein structure have been developed (Tu, 1986). Figure 4.4 is a simple representation of the structural features that have been determined from conventional Raman spectroscopy.

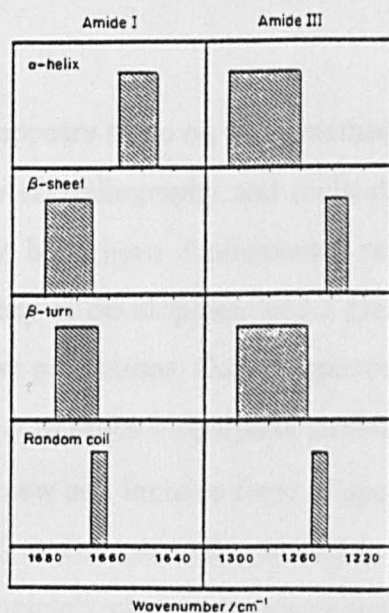


Figure 4.4 Different types of protein conformations from amide I and III vibrations



The precise positions of these bands have also been shown to be highly informative. For example, the exact position of the  $\alpha$ -helix Raman band in the  $\sim 900$  -  $1000\text{ cm}^{-1}$  region has been found to contain information about the geometry of the helix, particularly the  $\varphi$  angle (Sengupta & Krimm, 1987).

Fourier transform-infrared (FT-IR) spectroscopy has also been applied to the study of biopolymers (Mantsch *et al.*, 1986). Analysis of the amide I and amide III vibrational bands can lead to the detection of  $\alpha$ -helix,  $\beta$ -sheet and possibly tertiary structure. However, the deconvolution techniques that are required in FT-IR spectroscopy can sometimes produce misleading conclusions.

The technique of VCD has been used to develop empirical correlations of VCD spectral features with respect to the stereochemical aspects of proteins (Keiderling & Pancoska, 1993). These analyses have mainly been restricted to the amide I region providing secondary structure predictions. VCD appears to be 'blind' to tertiary structure which is therefore a limitation on the technique as will be shown later.

## Conclusion

Currently there appears to be no ideal method to determine the structure of proteins. Although X-ray crystallography and multi-dimensional NMR can provide complete structures they both have fundamental restrictions. The computational methods are in their infancy of development and a great deal of work is yet required to gain confidence in these predictions. Current spectroscopic techniques also appear to be limited in their ability to probe biopolymer structure.

Since ROA is a new and incisive form of spectroscopy (with both  $\text{H}_2\text{O}$  and  $\text{D}_2\text{O}$  being ideal solvents) that shows real potential in this area it warrants intensive studies. It is still not completely clear as to where the technique is going or what it can eventually elucidate in structural terms; but the following chapters will describe the recent findings of this project with respect to protein structure and dynamics.

## Chapter 5

### ROA of Peptides and Polypeptides

Polypeptide chains have long been known to adopt  $\alpha$ -helical,  $\beta$ -sheet and so-called unordered conformations under certain pH and temperature conditions (Greenfield & Fasman, 1969; Yu *et al.*, 1973; Itoh *et al.*, 1976; Painter & Koenig, 1976; Yu, 1977). They would therefore appear to be ideal starting points to determine structural features in a new form of biopolymer spectroscopy. The polypeptides to be discussed in this chapter are poly-L-lysine and poly-L-glutamic acid. A model  $\beta$ -turn peptide, L-Pro-L-Leu-Gly-NH<sub>2</sub> was also studied.

There has recently been renewed interest in unordered conformations in particular and trying to use these conformations to generate a "baseline" to calculate the amount of structure in other species. Polypeptide studies are of great importance because there have been attempts to extend the results to study protein folding pathways (Creighton, 1988). Even today it is not fully understood if the starting point of protein folding is the fully unfolded state or if there are regions with residual order. Protein folding will be discussed in Chapter 7.

Poly-L-glutamic acid adopts an  $\alpha$ -helical structure at low pH and  $\sim 20^\circ\text{C}$ , a random coil state at high pH and  $\sim 20^\circ\text{C}$ , and two different antiparallel  $\beta$ -sheet structures at high pH and in the high temperature regions of  $\sim 40 - 85^\circ\text{C}$  ( $\beta_1$ ) and  $>85^\circ\text{C}$  ( $\beta_2$ ), respectively. Poly-L-lysine adopts an  $\alpha$ -helical state at high pH and low temperature, a random coil state at low pH and  $\sim 20^\circ\text{C}$ , and a  $\beta$ -sheet state at high pH but high temperature. Varying the pH and temperature changes the ionization state of the side-groups therefore allowing different conformations to be adopted. The L-Pro-L-Leu-Gly-NH<sub>2</sub> peptide is known to form a model type II  $\beta$ -turn structure and is therefore also useful for understanding certain features in protein ROA spectra.

Although preliminary ROA spectra of these peptides and polypeptides have previously been published (Wen, 1992; Wen *et al.*, 1994) a more detailed study based on higher quality ROA data is presented here (Wilson *et al.*, 1996).

Although there is wide agreement that poly-L-lysine adopts an unordered (or random coil) state at low pH, the actual description of this unordered state has been at the centre of much controversy. A model proposed by Tiffany and Krimm (1968a; 1969) in the late 1960's was the first to question what the actual "random coil" conformation represented. Based on ECD results of the so-called random coil conformations of poly-L-lysine and poly-L-glutamic acid, Tiffany and Krimm suggested that some sort of structure is present. They deduced that the actual "random coil" conformation could be described by a helix model which contains left-handed helical segments of approximately 4 - 7 residues in length with 2.5 - 3 residues per turn which are linked by sharp bends. Their model was based on three simple arguments, two of which have now been revealed to have serious flaws.

To date their only uncontested observation involved the strong resemblance of the poly-L-proline II (PPII) CD spectra with that of the random coil ECD spectra (Krimm & Mark, 1968; Krimm *et al.*, 1969; Hiltner *et al.*, 1972) and this was one of the main observations on which Tiffany and Krimm based their model. Poly-L-proline is known to occur in two ordered structural forms: in the solid state it exists as a right-handed helix of cis imide groups with ten residues in three turns (PPI) and in solution it is a left-handed helix of trans imide groups with three residues in one turn (PPII). Poly-L-proline cannot form normal  $\alpha$ -helical or  $\beta$ -sheet structures because of its side-chain pyrrolidine ring and the lack of an amide N-H to allow hydrogen bonding to take place, as shown in Figure 5.1.

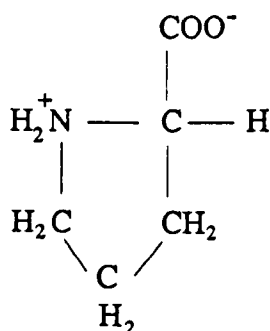


Figure 5.1 Illustration of the amino acid proline

Tiffany and Krimm's second argument involved the use of various polypeptides and proteins in concentrated salt solutions (sometimes after heating) as models for unordered polypeptides. Although these are unordered systems, their suitability as models at only moderate ionic strengths is questionable. Their model polypeptides were also subject to cis-trans isomerization effects which tended to cancel out any optical rotation. The rotational barrier for this isomerization was later found to be dependent on concentration and temperature, casting doubt on their conclusions (Piez & Sherman, 1970; Jenness *et al.*, 1976). Finally, their theoretical calculations also suggested that a polypeptide whose side-chains are charged would form an extended helical conformation, such as in the PPII helix. However, these calculations were undermined since a regular helical conformation had been assumed, the entropy variations in the backbone and side-chains had not been considered. Screening by counterions had also been neglected. Doubt was thus cast on this model and in particular Woody (1977) initially argued that the random coil conformation was unlikely to have well-defined structure despite the fact that under the experimental conditions of high charge density or at low temperature it was not unreasonable to assume that some regions of conformational space may be favoured over others.

VCD, which measures the differential absorption of right and left circularly polarized IR radiation, has provided valuable information on the random coil conformation. VCD is an incisive probe here because, unlike UV CD where there are two overlapping transitions, namely  $\Pi^* \leftarrow \Pi$  and  $\Pi^* \leftarrow n$  which are the source of the absorption in the 180 - 230 nm region, suitable VCD bands can be selected that just originate in one transition (e.g. namely the amide I region). The VCD signal is also relatively sharp with the bandwidth being about the same width as the exciton splitting. VCD results (Paterlini *et al.*, 1986; Birke *et al.*, 1991; Dukor & Keiderling, 1991) obtained for the random coil conformations of poly-L-lysine and poly-L-glutamic acid are found to be identical in shape but smaller in magnitude to the PPII spectra and therefore support the existence of some sort of left-handed helical structure. The same VCD was also obtained for the poly-L-lysine at a variety of low

pH's, which also suggests a single conformation and not a mixture of states in thermodynamic equilibrium (Dearborn & Wetlaufer, 1970). The general consensus of opinion, based mainly on UV CD and VCD results, now appears to be that the random coil conformation contains short stretches of left-handed three-fold helix.

It was therefore decided to perform a detailed study of these model peptides and polypeptides with a view to obtaining new information on the nature of any structure present and also with a view to the interpretation of protein ROA spectra.

## 5.1 Experimental

Samples of L-lysine, L-Pro-L-Leu-Gly-NH<sub>2</sub>, poly-L-lysine (hydrobromide,  $\bar{M}_w \approx 4,000, 26,000, 56,000$  and  $268,000$  by viscosity) and poly-L-glutamic acid (hydrobromide,  $\bar{M}_w \approx 20,000$  by viscosity) were obtained from Sigma, as was the D<sub>2</sub>O. Analytical grade glycine and NaCl from Sigma was used to prepare the buffered and high salt concentration samples. The L-lysine samples were dissolved in filtered deionised H<sub>2</sub>O (and D<sub>2</sub>O) to a concentration of 2M. The poly-L-lysine at pH 3 (and pD 3) and 100 mg/ml was obtained by using a 100 mM glycine buffer. The ~100 mg/ml poly-L-glutamic acid sample at pH 10 (and pD 10) was also prepared in a glycine buffer. Great difficulty was experienced in obtaining the poly-L-lysine solution at high pH and ~100 mg/ml. The procedure finally adopted was to add a 10% excess of NaOH to deprotonate all of the amide nitrogens in an ice bath. When the sample was allowed to warm up the  $\alpha$ -helical structure was found to crosslink, leading to the formation of a gel. Cold dry air was therefore blown over the sample cell in order to create an environment of  $-2^\circ\text{C}$  during the recording of the ROA spectra. The L-Pro-L-Leu-Gly-NH<sub>2</sub> amide was dissolved in H<sub>2</sub>O (and D<sub>2</sub>O) to a concentration of 0.3 M.

Unfortunately, due to gel formation at the high temperature required for the  $\beta$ -sheet formation of poly-L-glutamic acid and due to the insufficient solubility at low pH required for the  $\alpha$ -helical conformation, only the random coil conformation

of poly-L-glutamic acid obtained at pH 10 and 20°C (Doty *et al.*, 1957; Fasman, 1967, 1978) could be studied with ROA. Similarly, since the  $\beta$ -sheet form of the poly-L-lysine (pH 11, T = 50°C) also forms a gel its ROA could not be measured. Hence it was not possible to obtain a complete range of  $\alpha$ -helix,  $\beta$ -sheet and random coil conformations for either poly-L-lysine or poly-L-glutamic acid. The ROA of glutamic acid itself was also unobtainable due to insufficient solubility. It was also unfortunate that poly-L-proline is insufficiently soluble for ROA measurements.

All sample solutions were filtered through 0.22  $\mu$ m Millipore filters into quartz microfluorescence cells and allowed to equilibrate for at least one day at the temperature the ROA data was collected at. Samples were then exposed to the laser beam for a few hours to reduce fluorescence and to allow the sample to stabilise in the beam (a power of ~600 mW was present at the sample). ROA acquisition times were ~3 h for L-lysine , ~7 h for L-Pro-L-Leu-Gly-NH<sub>2</sub> and ~10 h for the polypeptides.

## 5.2 Results and Discussion

The backscattered Raman and ROA spectra of the samples under a variety of conditions will now be discussed.

### L-Lysine

The backscattered Raman and ROA spectra of the amino acid L-lysine (Figure 5.2) will first of all be discussed with the main emphasis on the CH<sub>2</sub> modes. L-lysine has four CH<sub>2</sub> side-groups as shown in Figure 5.3.

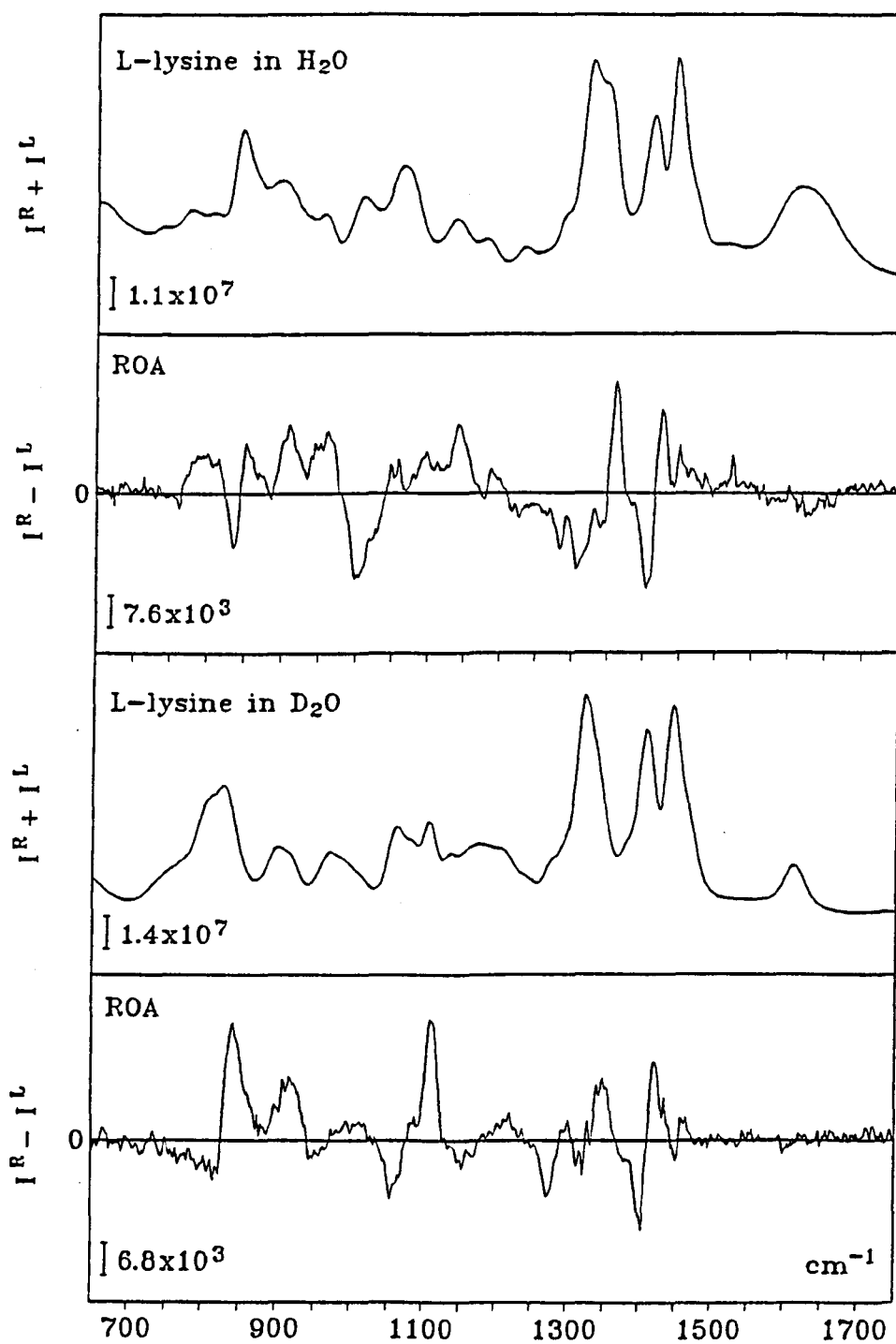


Figure 5.2 Backscattered Raman and ROA of L-lysine

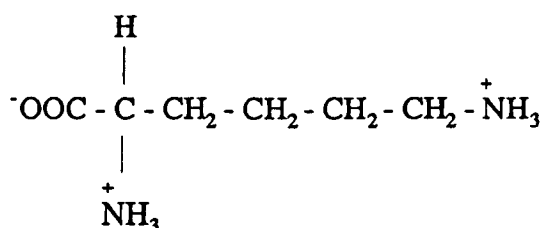


Figure 5.3 Lysine has a long basic side-chain

Since its relatively long side-chains may adopt some sort of conformation in the corresponding polypeptide, it is important to be able to decide whether signals observed in the polypeptide arise from  $\text{CH}_2$  side-groups or from the polypeptide secondary structure.

The ROA of L-lysine shows some similarities to that of L-alanine which therefore serves as a useful comparison since L-alanine has been the subject of a detailed experimental and *ab initio* ROA study (Barron *et al.*, 1991, 1992; Nafie *et al.*, 1994; Polavarapu & Deng, 1994) comparing it to other amino acids (Gargaro *et al.*, 1993). The ROA spectra of L-lysine and L-alanine are very similar in the  $\sim 700 - 1200 \text{ cm}^{-1}$  range, which can be attributed to normal modes involving vibrational coordinates of the  $-\text{CH}(\text{NH}_3^+)(\text{CO}_2^-)$  group. Most of the differences can be attributed to C-C stretch with  $\text{CH}_2$  rock (at the lower end of this wavenumber range) and  $\text{CH}_2$  twist (at the higher end) coordinates of the  $-(\text{CH}_2)_4-$  groups (Dollish *et al.*, 1974; Colthup *et al.*, 1975).

L-lysine in  $\text{H}_2\text{O}$  shows negative ROA in the  $\sim 1250 - 1340 \text{ cm}^{-1}$  range, which is not shown by L-alanine but is present in L-serine, where the negative ROA was attributed to  $\text{CH}_2$  twist and wag co-ordinates coupled with COH bends (Gargaro *et al.*, 1993). In the case of the L-lysine the four  $\text{CH}_2$  groups are responsible for the negative ROA, with a significant contribution from NH deformations suggested by the changes in L-lysine in  $\text{D}_2\text{O}$  observed in this region.

In  $\text{H}_2\text{O}$  the ROA of L-lysine has a large positive band at  $1363 \text{ cm}^{-1}$  and a large negative band at  $1405 \text{ cm}^{-1}$ . This is also a common feature of L-alanine, L-serine and L-threonine which has previously been assigned to  $\text{C}_\alpha\text{-H}$  deformations and to the symmetric  $\text{CO}_2^-$  stretch and therefore L-lysine can be similarly assigned. The



positive band at  $1445\text{ cm}^{-1}$  in the ROA of L-lysine can be assigned to the  $\text{CH}_2$  scissors deformation from previous conventional Raman work (Dollish *et al.*, 1974; Colthup *et al.*, 1975). The ROA of L-lysine in  $\text{D}_2\text{O}$  has a very similar appearance to that in  $\text{H}_2\text{O}$  in this region which provides confidence for the assignments of these bands.

### **$\alpha$ -Helical Poly-L-Lysine ( $\bar{M}_w \approx 26,000$ )**

The backscattered Raman and ROA spectra of  $\alpha$ -helical poly-L-lysine were measured with samples having four different molecular weights:  $\bar{M}_w \approx 4,000$ , 26,000, 56,000 and 268,000. The poly-L-lysine with  $\bar{M}_w \approx 26,000$  will be discussed first and then these results related to the other samples.

Figure 5.4 shows the backscattered Raman and ROA spectra of poly-L-lysine with  $\bar{M}_w \approx 26,000$  in an  $\alpha$ -helical conformation. It is reassuring that the Raman spectra are virtually identical to previously published spectra (Yu *et al.*, 1973; Painter & Koenig, 1976). The absence of significant Raman intensity in the  $\sim 1200 - 1260\text{ cm}^{-1}$  region in the  $\text{H}_2\text{O}$  spectrum implies that the sample is in an  $\alpha$ -helical state and that effective cooling has been maintained throughout the experiment. Raman intensity in this region is indicative of random coil and  $\beta$ -sheet conformations (Yu *et al.*, 1973). It is also obvious at a glance that the ROA spectra of L-lysine and poly-L-lysine are completely different, suggesting that most of the ROA is due to vibrations in the peptide backbone rather than side-groups (or that the  $\text{CH}_2$  side-groups have completely different conformations in the amino acid and polypeptide).

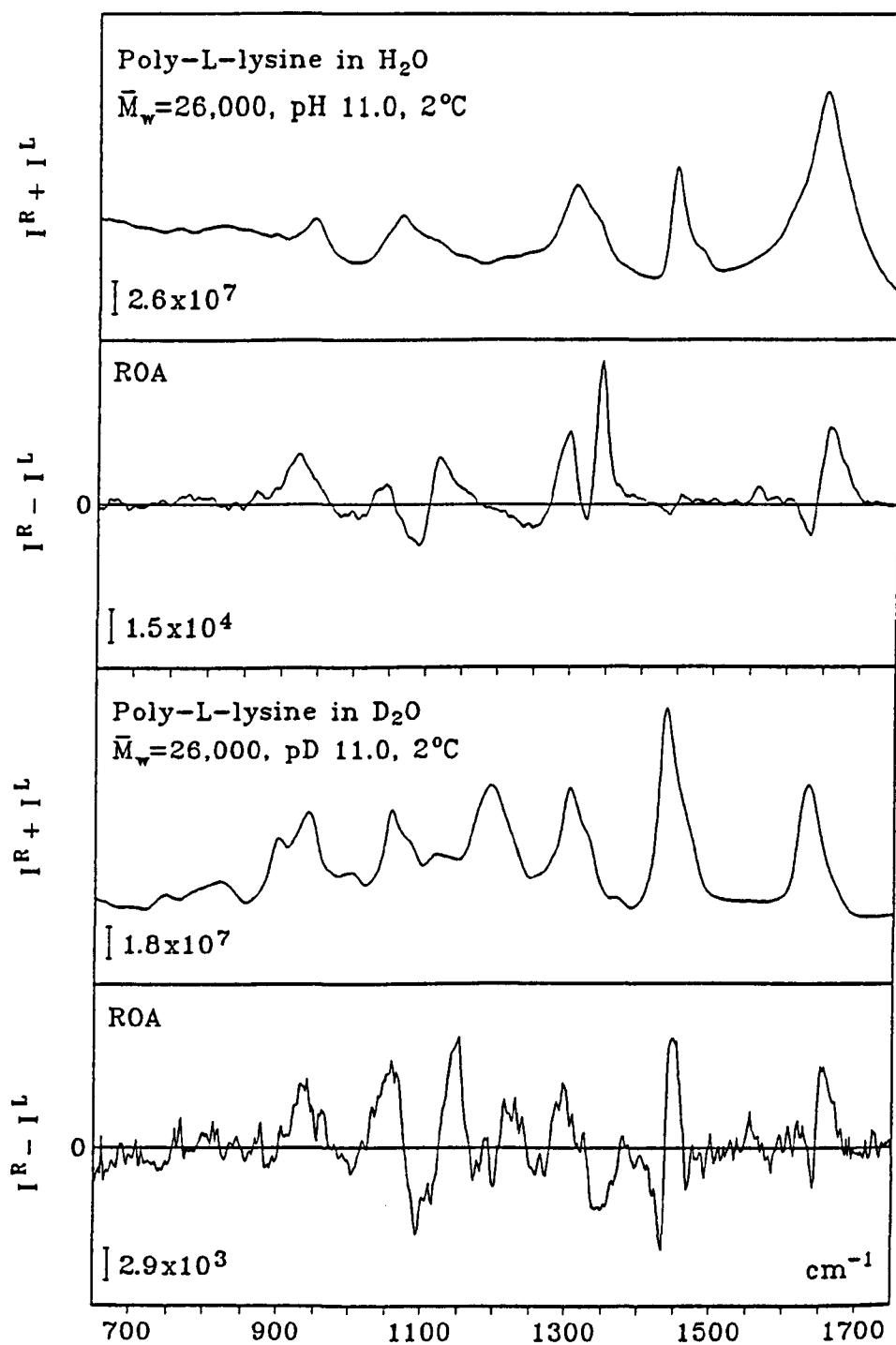


Figure 5.4 Backscattered Raman and ROA of  $\alpha$ -helical poly-L-lysine ( $\bar{M}_w \approx 26,000$ )

In the H<sub>2</sub>O ROA spectrum there is a broad positive ROA band at ~915 - 973 cm<sup>-1</sup> peaking at 931 cm<sup>-1</sup>. Conventional Raman bands at 943 cm<sup>-1</sup> and 1068 cm<sup>-1</sup> have previously been assigned to backbone and side-chain skeletal C<sub>α</sub>-C stretch modes of  $\alpha$ -helical structures (Yu, 1977; Tu, 1986). This feature has also been observed in early ROA spectra of  $\alpha$ -helical proteins such as bovine serum albumin (BSA) and insulin (Wen *et al.*, 1994a), which will be discussed in greater detail in Chapter 6. It is also reassuring that in  $\beta$ -sheet proteins such as  $\alpha$ -chymotrypsin, preliminary ROA studies have shown these signals to be absent (Wen *et al.*, 1994b).  $\alpha$ -helical poly-L-lysine also has a negative ROA band at 1000 cm<sup>-1</sup> which is also present but always appearing smaller in  $\alpha$ -helical proteins. L-lysine also has negative ROA in this region which might suggest that the negative ROA is due to both  $\alpha$ -helix modes and side-groups in the poly-L-lysine. A similar method of reasoning can be used for the positive ROA band at 1046 cm<sup>-1</sup> because this band is not characteristic of  $\alpha$ -helical proteins and therefore may well be associated with side-groups. However, the conservative ROA couplet centred at 1103 cm<sup>-1</sup>, negative at low wavenumber and positive at high wavenumber is a clear well defined  $\alpha$ -helix signature and is prominent in most  $\alpha$ -helical proteins (Wen *et al.*, 1994a). The  $\alpha$ -helical poly-L-lysine in D<sub>2</sub>O in the range ~900 - 1170 cm<sup>-1</sup> is very similar to that of the H<sub>2</sub>O spectrum, which is what would be expected since most of the intensity comes from backbone skeletal stretch modes. The slight changes that do occur are due to deuterium exchange of the N-H groups. A perplexing aspect, however, is that there is also a large drop in intensity (~5 fold) in the ROA D<sub>2</sub>O spectrum, for which there is no obvious explanation.

The next region of interest is the extended amide III which contains negative ROA intensity at lower wavenumber and positive at higher and is a common feature of almost all polypeptides and proteins studied to date. However, the differences that do occur are highly representative of the secondary structure and the loops and turns that are present (Barron *et al.*, 1994, 1996; Wen *et al.*, 1994a,b; Ford *et al.*, 1995; Wilson *et al.*, 1995; Wilson *et al.*, 1996). The amide III Raman band occurs at 1360 cm<sup>-1</sup> with two shoulders at 1334 cm<sup>-1</sup> and 1390 cm<sup>-1</sup>, which are assigned to the CH<sub>2</sub> twisting and CH<sub>2</sub> wagging vibrations, respectively. There is little change of these Raman bands in the deuterated sample. The sharp positive ROA peak at 1300 cm<sup>-1</sup> coupled with the small negative peak at 1265 cm<sup>-1</sup> is also observed

in  $\alpha$ -helical proteins and appears to be a good  $\alpha$ -helix signature (Wen *et al.*, 1994a; Ford *et al.*, 1995; Wilson *et al.*, 1995; Barron *et al.*, 1996). There is a sharp positive ROA band at  $1340\text{ cm}^{-1}$ , which vanishes in the  $\text{D}_2\text{O}$  spectra. This confirms its assignment to amide III-type coupled N-H and  $\text{C}_\alpha$ -H deformations rather than  $\text{CH}_2$  deformations from the side-chains. Proteins such as BSA also show a sharp positive ROA band in this region (Barron *et al.*, 1994), which is not observed in  $\beta$ -sheet proteins (Wen *et al.*, 1994b) or in insulin (Wen *et al.*, 1994a). L-lysine has negative ROA in this region which again suggests that side-chain interactions are not involved. In preliminary BSA spectra it was suggested that this band could be due to rigid loop structure with a particular type of local order (Barron *et al.*, 1994, 1996; Wen *et al.*, 1994a). It is also of interest that the ROA of fully deuterated lysozyme is similar to the  $\text{D}_2\text{O}$  solution in the amide III region. This can be explained by the complete loss of the N-H deformations, to the associated normal modes in this region, with the normal modes now being dominated by  $\text{C}_\alpha$ -H deformations (Ford *et al.*, 1994, 1995). ROA is thus providing direct evidence that poly-L-lysine in its supposed  $\alpha$ -helical conformation contains stretches of  $\alpha$ -helix connected by rigid loop sections with the same type of local order as many of the loops in BSA. The origin and importance of the  $1340\text{ cm}^{-1}$  ROA band (which is now thought to be  $3_{10}$ -helix) will be discussed again later in this section and in later chapters.

The strong sharp Raman band at  $1444\text{ cm}^{-1}$  due to side-group  $\text{CH}_2$  scissors deformation surprisingly has little ROA, even though L-lysine shows large ROA in this region. The  $\text{D}_2\text{O}$  sample on the other hand has a couplet centred at  $1443\text{ cm}^{-1}$ , negative at low wavenumber and positive at high. This couplet is due to the  $\text{CH}_2$  scissors deformation coupling with the amide II mode that has shifted to lower wavenumber. This is called the amide II' region and the significant ROA that it shows in peptides and proteins has been attributed to the coupling of N-D deformations with side-group  $\text{CH}_2$  or  $\text{CH}_3$  deformations (Ford *et al.*, 1994, 1995).

In the amide I region there is a positively biased ROA couplet with a crossover point at  $1642\text{ cm}^{-1}$ . Unlike VCD, ROA in the amide I region is not particularly sensitive to polypeptide and protein structure (Keiderling & Pancoska, 1993). A positively biased ROA couplet is also observed in insulin (Wen *et al.*, 1994a) and BSA (Barron *et al.*, 1995), which is to be expected since insulin has 61%  $\alpha$ -helix (Lee *et al.*, 1990) and human serum albumin has 67% helix (He & Carter,

1992). On deuteration the amide I Raman band shifts to a lower frequency by about  $16\text{ cm}^{-1}$ , caused by the loss of the N-H in-plane contribution to the normal mode.

### Unordered Poly-L-Lysine ( $\bar{M}_w \approx 26,000$ ) and Unordered Poly-L-Glutamic Acid ( $\bar{M}_w \approx 20,000$ )

The backscattered Raman and ROA spectra of unordered poly-L-lysine ( $\bar{M}_w \approx 26,000$ ) and poly-L-glutamic acid ( $\bar{M}_w \approx 20,000$ ) are shown in Figure 5.5 and Figure 5.6, where the conventional Raman spectra are again virtually identical to those in the literature (Yu *et al.*, 1973; Painter *et al.*, 1976). The backscattered Raman and ROA spectra of the unordered poly-L-lysine ( $\bar{M}_w \approx 26,000$ ) in 5M NaCl at  $20^\circ\text{C}$  and  $50^\circ\text{C}$  is shown in Figure 5.7. The unordered poly-L-lysine will first of all be discussed.

It can be seen that unordered poly-L-lysine in  $\text{H}_2\text{O}$  shows significant ROA structure throughout the spectrum with a positively-biased, highly structured, amide III couplet. This is in direct contrast to the ROA of an unfolded protein (see later) where there is a broad relatively unstructured conservative couplet in the extended amide III region with little ROA elsewhere (Ford *et al.*, 1995; Barron *et al.*, 1996).

An important ROA feature is the negative signal between  $\sim 900 - 971\text{ cm}^{-1}$ . Since this has the opposite sign to the corresponding signal in the  $\alpha$ -helical poly-L-lysine which has been suggested above to be characteristic of right-handed  $\alpha$ -helix, it could well originate in left-handed helical structure. This negative band is also present but not so clear in the deuterated spectrum, with only a slight loss in intensity. Notice also that there is no positive ROA band at  $1340\text{ cm}^{-1}$ . It can thus be concluded that the regions connecting the left-handed helix do not have the local characteristics of  $3_{10}$ -helix, as found in the  $\alpha$ -helical polypeptide. This observation provides new support for the original proposal of Tiffany and Krimm (1968a) when they suggested that unordered homopolypeptides contain locally-ordered stretches of left-handed helix, as found in PPII. Recent VCD studies have also agreed with this conclusion (Paterlini *et al.*, 1986; Dubar *et al.*, 1991; Birke *et al.*, 1992). However, in addition the ROA data is suggesting that the stretches of left-handed helix are linked by other well-defined structures.

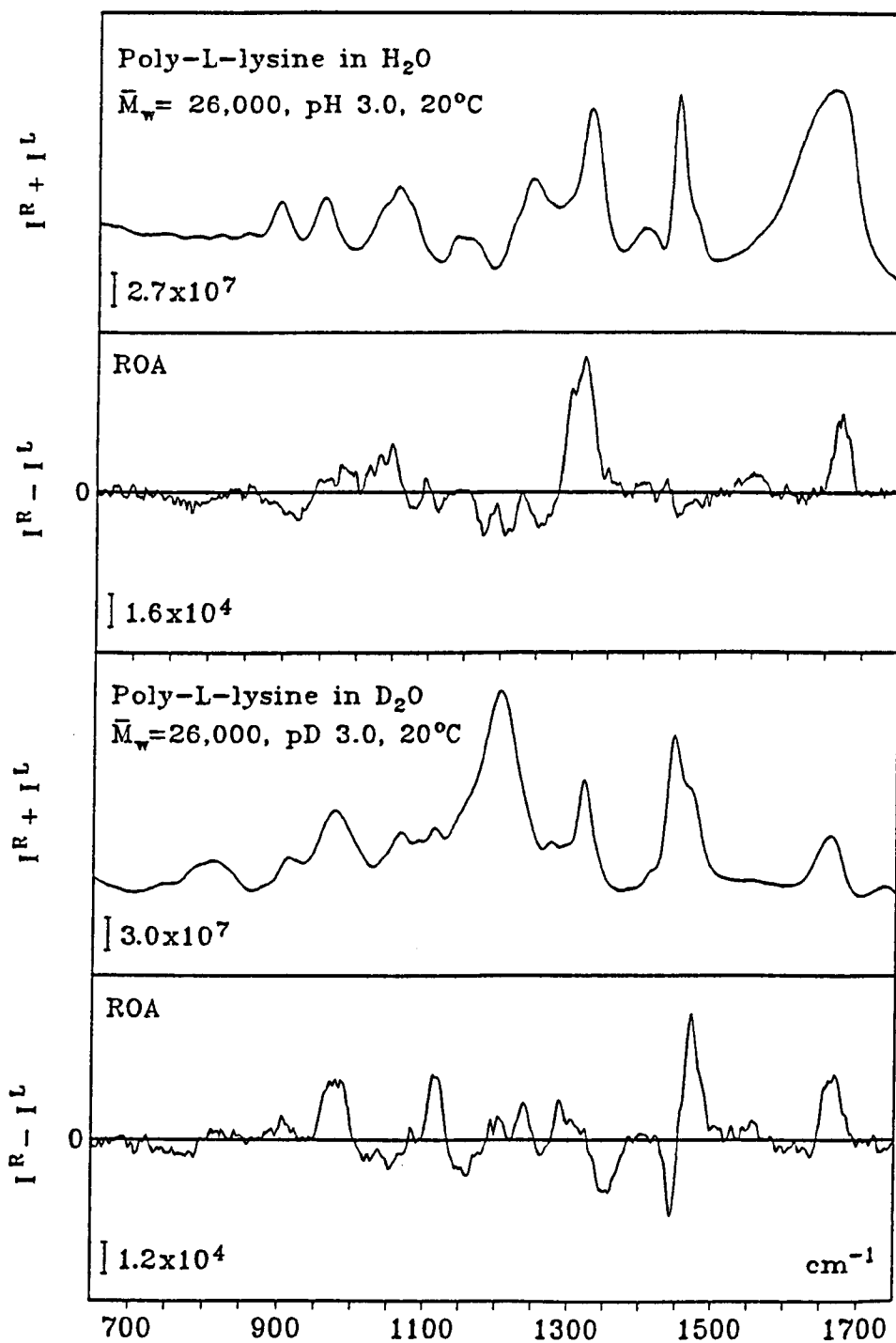


Figure 5.5 Backscattered Raman and ROA of unordered poly-L-lysine ( $\bar{M}_w \approx 26,000$ )

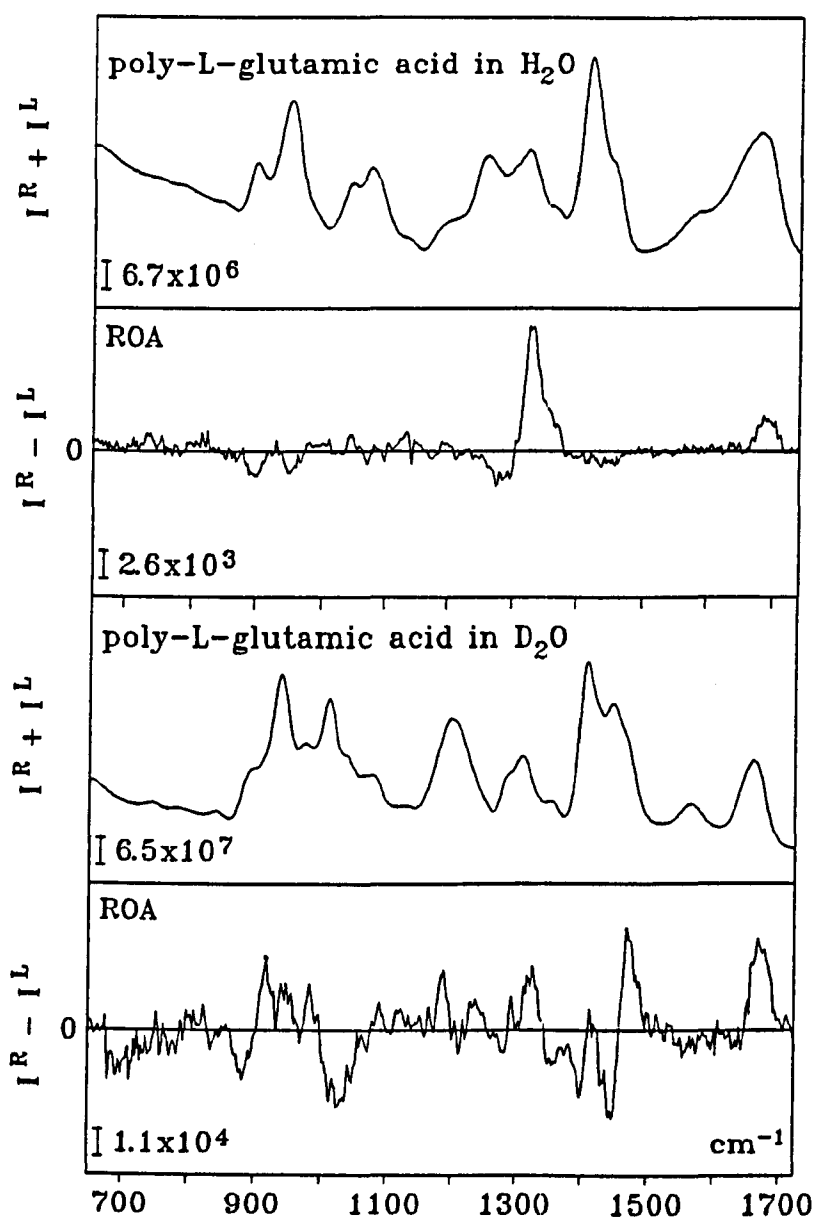


Figure 5.6 Backscattered Raman and ROA of unordered poly-L-glutamic acid ( $\bar{M}_w \approx 20,000$ )

The sharp positive ROA band at  $1314\text{ cm}^{-1}$  is characteristic of  $\beta$ -sheet proteins (see later) and provides strong evidence that there is a considerable amount of  $\beta$ -strands present. However, due to the lack of negative ROA in the range  $\sim 1340 - 1380\text{ cm}^{-1}$  which is characteristic of  $\beta$ -turns (see later) it can be suggested that there are no tight turns connecting the strands. This is perhaps not surprising since glycine and proline residues are usually required to accommodate tight turns in adjacent  $\beta$ -strands (Creighton, 1993). The broad ROA band in the range  $\sim 1000 - 1060\text{ cm}^{-1}$  also implies that  $\beta$ -sheet is present; this band is also common in some  $\beta$ -sheet proteins (see later). It is not yet fully understood if this is generated from sheet, strand, or simply individual residues with high  $\beta$ -structure  $\phi$  and  $\psi$  propensities.

The positive-negative ROA couplet with the positive part peaking at  $1048\text{ cm}^{-1}$  and negative part peaking at  $1075\text{ cm}^{-1}$  has recently been tentatively assigned to right-twisted  $\beta$ -strands. A similar couplet is observed in proteins with highly twisted double-stranded  $\beta$ -sheets such as ribonuclease A (there is also sometimes a negative ROA region between  $\sim 950 - 1000\text{ cm}^{-1}$  in proteins making this negative-positive-negative pattern an indication of the presence of right-twisted  $\beta$ -strands). This assignment is reinforced by the fact that proteins which contain relatively flat multi-stranded  $\beta$ -sheets do not contain this ROA pattern. An example of this is concanavalin A (see later) where the flat  $\beta$ -sheet exists in a  $\beta$ -barrel. This ROA pattern may therefore be a signature of the right-handed twist characteristic of the strand in a  $\beta$ -sheet structure that is actually equivalent to a left-handed helical conformation of the peptide backbone (Sengupta & Krimm, 1987; MacCallum *et al.*, 1995). The assignment of this ROA pattern to right-twisted  $\beta$ -strands will be explained in greater detail in Chapter 6.

Associated with the Raman band at  $1444\text{ cm}^{-1}$  in  $\text{H}_2\text{O}$ , there is a large associated negative ROA region accompanied by a positive region peaking at  $1550\text{ cm}^{-1}$ . This suggests that there is some sort of coupling mechanism between the  $\text{CH}_2$  deformations and the amide II modes.

In the amide I region the positive part of the ROA couplet is very similar to that in  $\alpha$ -helical poly-L-lysine. However, the negative part has almost completely disappeared. It should also be noted that the amide I in unordered poly-L-lysine is different to that found in unfolded (reduced) lysozyme, since it is shifted to higher wavenumber by  $10\text{ cm}^{-1}$  in the protein (Tu, 1986).



The amide III couplet which has its positive part peaking at  $1314\text{ cm}^{-1}$  and negative part at  $1265\text{ cm}^{-1}$  is interesting because there is a subsidiary positive peak at  $1300\text{ cm}^{-1}$ . In the high pH sample of poly-L-lysine this band was assigned to  $\alpha$ -helix, so there could also be sections of  $\alpha$ -helix present in the low pH sample. Although, the suggested left-handed helix might have been expected to have opposite amide III ROA sign patterns to that characteristic of the right-handed helix which would have cancelled out the signature from the  $\alpha$ -helix; however, it must be noted that interchain hydrogen bands are absent from the PPII helix (Dukar & Keiderling, 1991; Creighton, 1993) and that interchain hydrogen bands will also be absent in a left-handed helix. The left-handed conformation will be stabilised by either hydrogen bonding to water or to side-chains. This interchain hydrogen bonding of N-H groups will have a significant contribution to the normal modes in the extended amide III region, so perhaps it is not surprising that we do not get a mirror-image version of a right-handed helix in the amide III region. It might however be possible that there is no extended  $\alpha$ -helix and that the  $1300\text{ cm}^{-1}$  peak comes from individual residues with  $\alpha$ -helix  $\varphi$  and  $\psi$  propensities.

Due to the fact that N-H deformations contribute less in the backbone skeletal stretch region there is actually a couplet of opposite sign centred at  $1103\text{ cm}^{-1}$  to that found in the right-handed helix which might originate in left-handed helix. If the N-H groups had a significant contribution this change of sign would not be observed.

It is surprising that in the skeletal stretch region the ROA spectra in  $\text{H}_2\text{O}$  and  $\text{D}_2\text{O}$  are quite different. This is in sharp contrast to the  $\alpha$ -helical case where the  $\text{H}_2\text{O}$  and  $\text{D}_2\text{O}$  spectra are very similar at low wavenumber (although there is a dramatic loss in the intensity for the  $\text{D}_2\text{O}$  spectra). This may well have something to do with the deuteration of the N-H groups. The large couplet in  $\text{D}_2\text{O}$  centred at  $1130\text{ cm}^{-1}$ , positive at low wavenumber and negative at high, is of opposite sign to that in  $\alpha$ -helical poly-L-lysine, suggesting that this is a signature for left-handed helix. Another ROA couplet centred at  $1331\text{ cm}^{-1}$ , positive at low wavenumber and negative at high, is also present in  $\alpha$ -helical poly-L-lysine in  $\text{D}_2\text{O}$  and is most likely due to  $\text{C}_\alpha$ -H local deformations. An ROA couplet negative at low wavenumber and positive at high, centred at  $1454\text{ cm}^{-1}$  is found in the  $\text{CH}_2$  scissors deformation region. Although a similar couplet is found in  $\alpha$ -helical poly-L-lysine in  $\text{D}_2\text{O}$ , it is

shifted by  $20\text{ cm}^{-1}$  to a higher wavenumber due to different normal modes. This is probably due to the coupling between the amide II' and  $\text{CH}_2$  scissors deformation. The unordered poly-L-lysine in  $\text{D}_2\text{O}$  is very similar to that of the  $\text{H}_2\text{O}$  sample in the amide I region.

Recent VCD experiments (Paterlini *et al.*, 1986) on unordered poly-L-lysine reported the disappearance of almost all of the amide I VCD on going to 5M NaCl solution. Due to this interesting observation it was decided to measure the ROA of poly-L-lysine under these conditions at  $20^\circ\text{C}$  and  $50^\circ\text{C}$ . These backscattered Raman and ROA spectra are presented in Figure 5.7. The results are very interesting because, unlike the VCD spectra, the ROA spectra are full of structure. It can thus be concluded that there is still a considerable amount of ordered structure present at  $20^\circ\text{C}$ . The left-handed helical signature indicated by a negative ROA band peaking at  $926\text{ cm}^{-1}$  is however much weaker with the added 5M NaCl. It is also interesting that features ascribed to  $\alpha$ -helix in the backbone skeletal stretch, amide III and amide I regions have increased; while the positive peak at  $1314\text{ cm}^{-1}$  which is due to  $\beta$ -sheet is unchanged. This effect is also emphasized by the appearance of the negative-positive ROA couplet at  $1103\text{ cm}^{-1}$ , which is an  $\alpha$ -helix signature and reinforces the idea that in the unordered polypeptide the signal is being cancelled out by left-handed helical structure. The broad negative-positive couplet with its negative part peaking at  $1448\text{ cm}^{-1}$  and positive part at  $1554\text{ cm}^{-1}$ , is much larger in unordered 5M NaCl sample than in the unordered sample and not even present in the  $\alpha$ -helical sample is due to coupling between side-group  $\text{CH}_2$  deformations and the backbone amide II. In the  $50^\circ\text{C}$  sample the trend continues with increasing  $\alpha$ -helix and the loss of left-handed helix.

The conclusion that can be drawn from these observations is that left-handed helical, right-handed helical and  $\beta$ -strand structures are present in unordered poly-L-lysine, with right-handed helix being favoured under destabilising conditions. This result fits with the conclusions from temperature-dependent UV CD studies of poly-L-lysine (Drake *et al.*, 1988). Their hypothesis was that over a range of  $-100^\circ\text{C}$  to  $+80^\circ\text{C}$ , structures containing left-handed helix became more favoured at low temperatures (Woody, 1992). However, their observations were not able to provide much information about the nature of the high temperature structures.

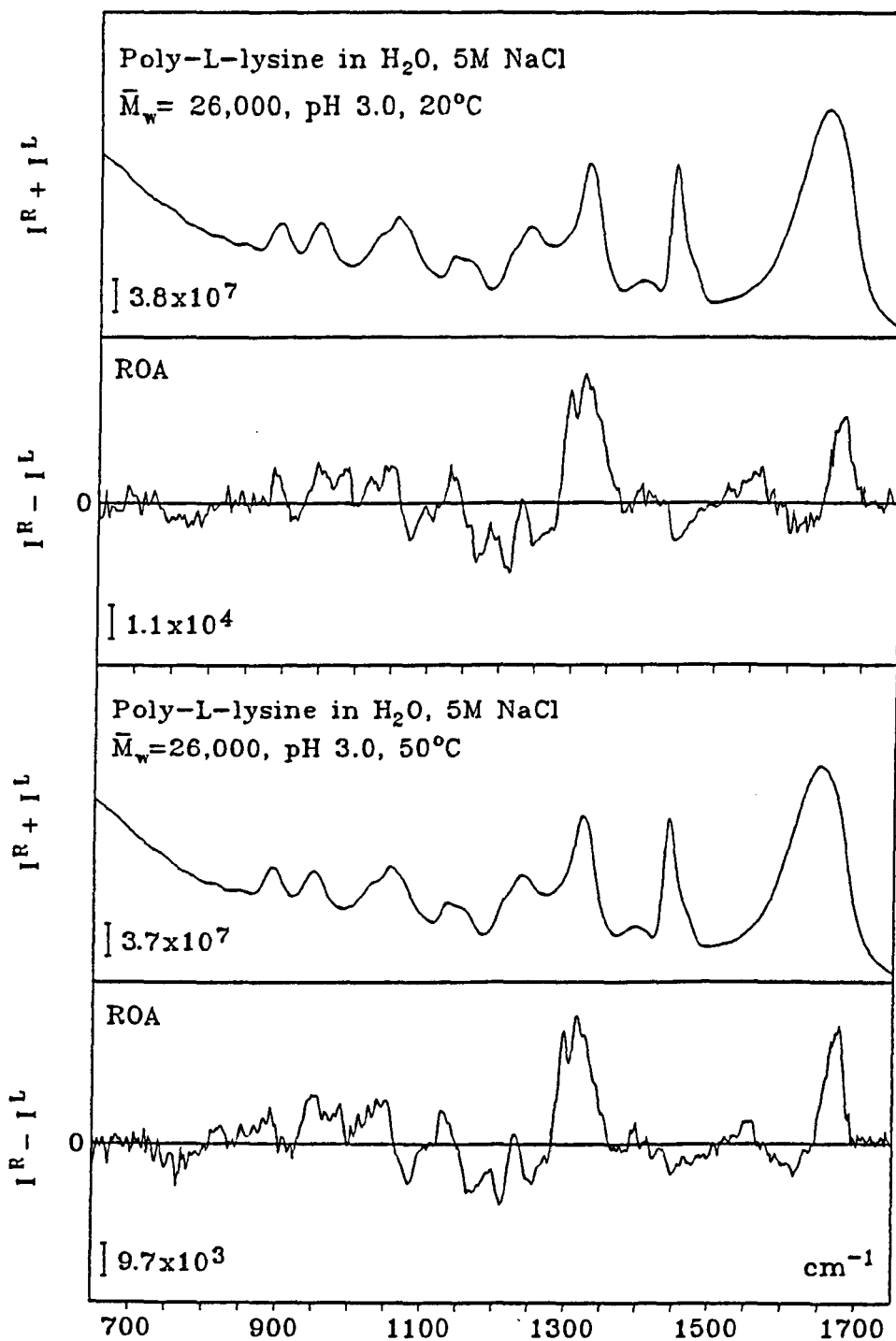


Figure 5.7 Backscattered Raman and ROA of unordered poly-L-lysine ( $\bar{M}_w \approx 26,000$ ) in 5M NaCl

**Poly-L-Glutamic Acid**

In Figure 5.6 the backscattered Raman and ROA spectra of random coil poly-L-glutamic acid are presented. Although generally similar to the random coil form of poly-L-lysine there are some differences in the detailed structure. Significantly there is only one negative ROA band in the  $\sim 1235 - 1305 \text{ cm}^{-1}$  region of the amide III in the poly-L-glutamic acid as compared to three in the random coil poly-L-lysine. The random coil poly-L-glutamic acid appears to be more disordered than the random coil poly-L-lysine, which probably has something to do with the side-groups.

**Poly-L-Lysine with Different Molecular Weights**

An analysis of  $\alpha$ -helical poly-L-lysine samples with  $\bar{M}_w \approx 4,000, 26,000, 56,000$  and  $268,000$  provides an excellent example of the sensitivity of ROA to subtle changes of backbone conformation. These spectra are shown in Figure 5.8 - 5.10.

L-lysine has a molecular weight of 147 and therefore it is relevant to calculate the approximate number of residues in each weight of poly-L-lysine.

Mol. wt. of poly-L-lysine	No. of residues in chain
4,000	27
26,000	177
56,000	380
268,000	1823

Table 5.1 Number of L-lysine units in poly-L-lysine samples of different molecular weight.

The striking aspect of the differences in the spectra is highlighted by monitoring the  $1340\text{ cm}^{-1}$  ROA peak (which has been deduced to be  $3_{10}$ -helix as will be explained in Chapter 6). The  $\alpha$ -helical poly-L-lysine of  $\bar{M}_w \approx 268,000$  shows a remarkable increase by a factor of  $\sim 1.5$  in the intensity of the  $1340\text{ cm}^{-1}$  positive ROA band compared to the  $\bar{M}_w \approx 26,000$  sample. The  $\bar{M}_w \approx 56,000$  sample is between these two spectra, indicating that the changes occur gradually on increasing molecular weight. It should also be noted that the Raman spectra for the different molecular weights are virtually identical, emphasising the greatly enhanced sensitivity of ROA to subtle conformational features.

Looking at the  $\bar{M}_w \approx 268,000$  sample the  $\alpha$ -helix band at  $931\text{ cm}^{-1}$  has broadened and shifted to lower wavenumber, which might be associated with the length of the  $\alpha$ -helix segments that are left after conversion to  $3_{10}$ -helix.

Standard techniques such as UV CD and NMR have great difficulty in differentiating between  $3_{10}$ - and  $\alpha$ -helix (Millhauser, 1995). VCD can distinguish the two but no aqueous solution results have been reported (Yasui *et al.*, 1986a,b; Yoder *et al.*, 1995).

A FTIR study of  $\alpha$ -helical poly-L-lysine revealed a band at  $1638\text{ cm}^{-1}$ , which is significantly lower than that for  $\alpha$ -helical structures in proteins (Jackson *et al.*, 1989) and close to the FTIR maximum reported for the  $3_{10}$ -helical peptides and to that assigned to  $3_{10}$ -helix in proteins (Prestrelski *et al.*, 1991). Jackson *et al.* (1989) also stated that additional amide I bands were observed as the molecular weight of the poly-L-lysine was increased, and suggested that these bands might arise "from helical structures of slightly different hydrogen-bonding characteristics possibly due to disruption of the hydrogen-bonding in regions of the chain which become folded or twisted as the chain becomes longer" (Jackson *et al.*, 1989). The increasing intensity of the  $3_{10}$ -helix content as the molecular weight is increased may be due to the greater tangling with many more sharp turns which the more regular  $\alpha$ -helix cannot support.

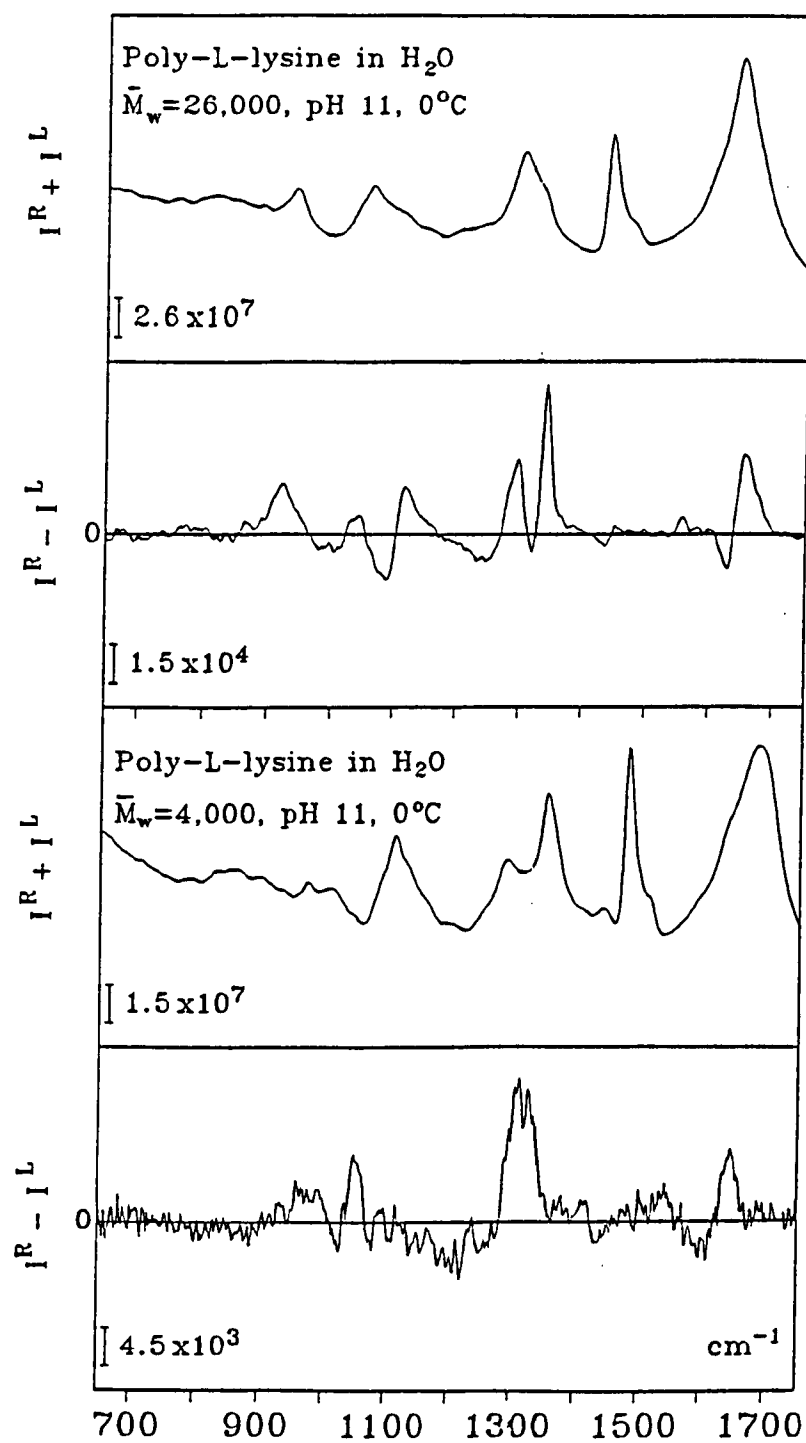


Figure 5.8 Backscattered Raman and ROA of  $\alpha$ -helical poly-L-lysine ( $\bar{M}_w \approx 26,000, 4,000$ )

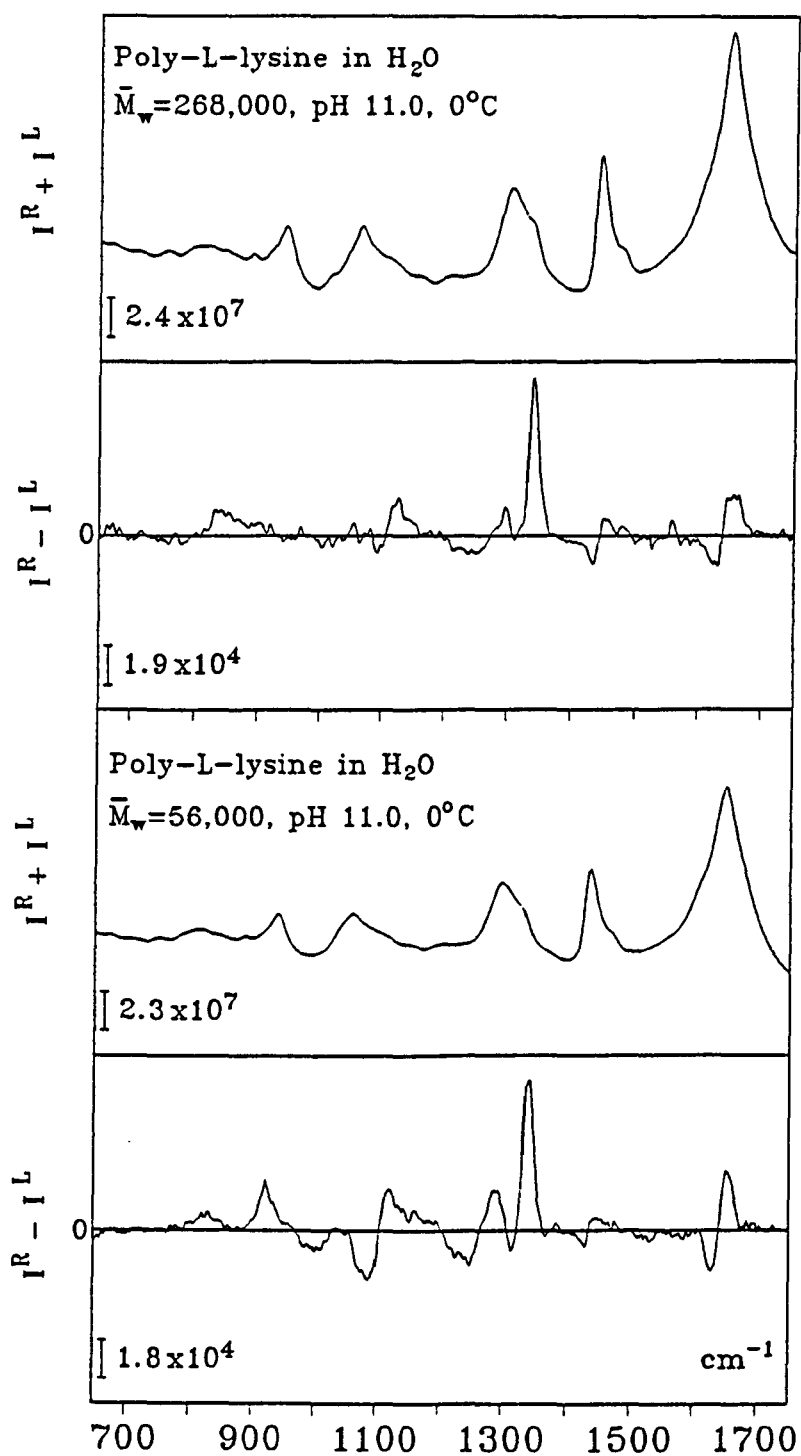


Figure 5.9 Backscattered Raman and ROA of  $\alpha$ -helical poly-L-lysine ( $\bar{M}_w \approx 268,000, 56,000$ )

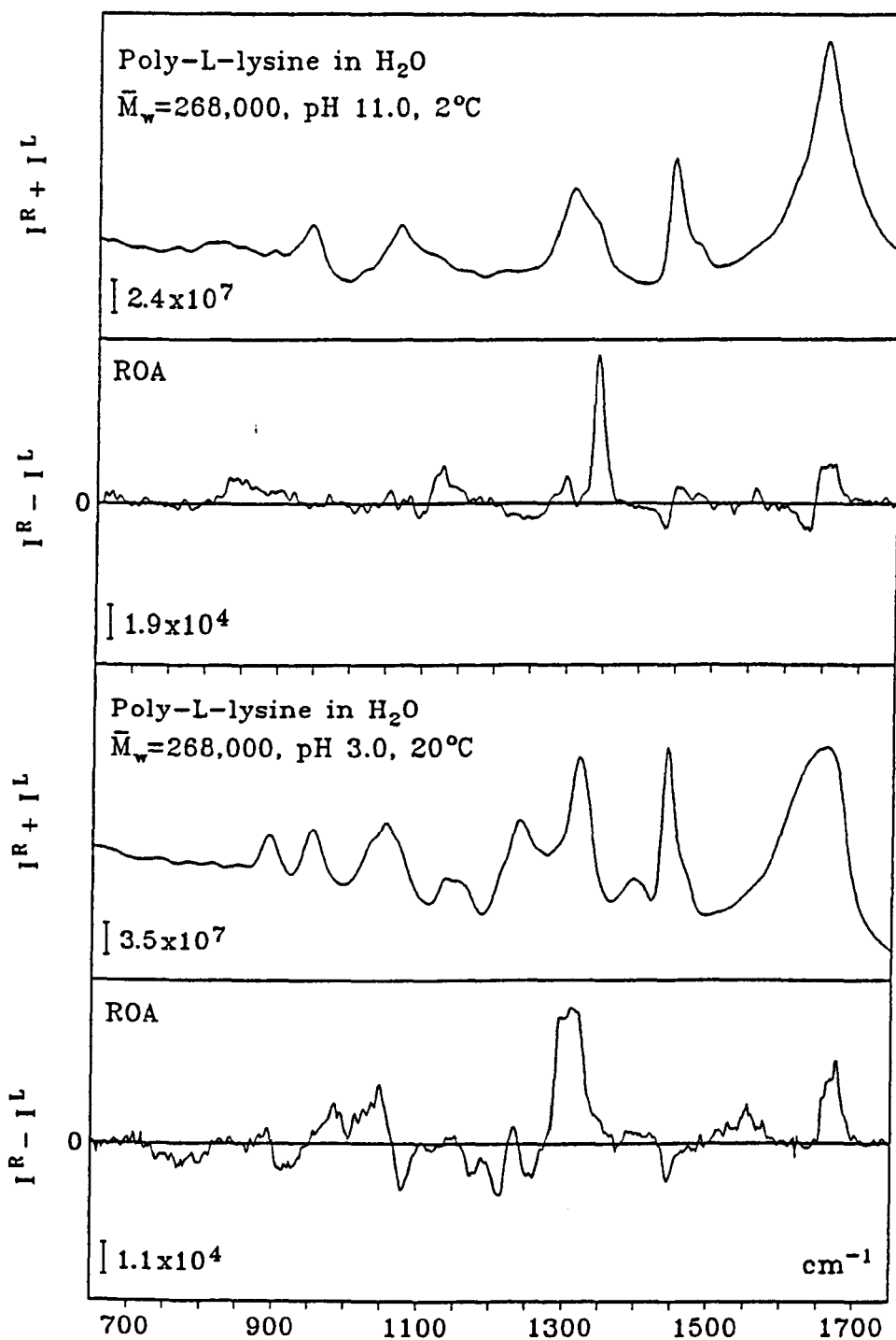


Figure 5.10 Backscattered Raman and ROA of unordered poly-L-lysine and  $\alpha$ -helical poly-L-lysine ( $\bar{M}_w \approx 268,000$ )



The backscattered Raman and ROA spectra of poly-L-lysine of  $\bar{M}_w \approx 4,000$  are shown in Figure 5.8. The poly-L-lysine of  $\bar{M}_w \approx 4,000$  appears to be a superposition of the helix and random coil spectra.

The ROA of the unordered poly-L-lysine of  $\bar{M}_w \approx 268,000$  is shown in the bottom half of Figure 5.10. The ROA spectrum is similar to the  $\bar{M}_w \approx 26,000$  form with many of the bands now sharper and stronger. However, a major difference is a large increase in the negative band at  $1075\text{ cm}^{-1}$ , which now forms a strong couplet with the positive part peaking at  $1048\text{ cm}^{-1}$ . A similar couplet is also observed in the  $\bar{M}_w \approx 26,000$  unordered sample with high salt concentration. The  $\bar{M}_w \approx 268,000$  random coil poly-L-lysine may therefore have a greater content of right-twisted  $\beta$ -strand as compared to the lower molecular weight form.

**L-Pro-L-Leu-Gly-NH<sub>2</sub>**

In Figure 5.11 the backscattered Raman and ROA spectra in H<sub>2</sub>O and D<sub>2</sub>O of the tripeptide L-Pro-L-Leu-Gly-NH<sub>2</sub> are shown. The tripeptide is known to exist in a type II  $\beta$ -turn conformation in the crystal from X-ray diffraction (Reed & Johnston, 1973) and from NMR in dimethylsulphoxide solution (Higashijima *et al.*, 1978). An illustration of the type II  $\beta$ -turn structure is shown in Figure 5.12.

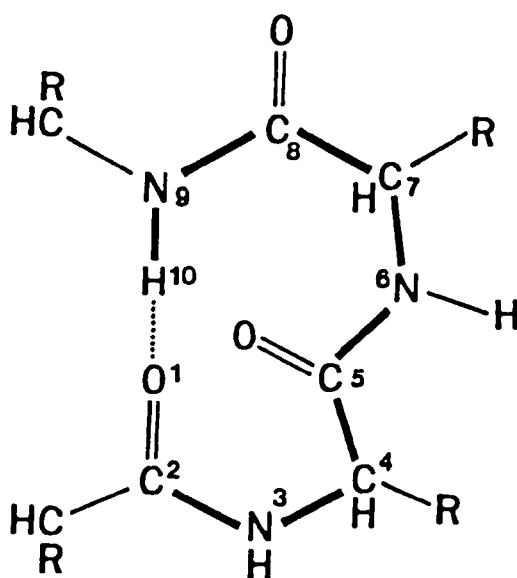


Figure 5.12 A model of the type II  $\beta$ -turn structure

It can be seen that the conformation is a rigid ten-membered turn structure, stabilized by an intramolecular hydrogen bond between the carbonyl of the proline residue and a proton of the terminal amide group.

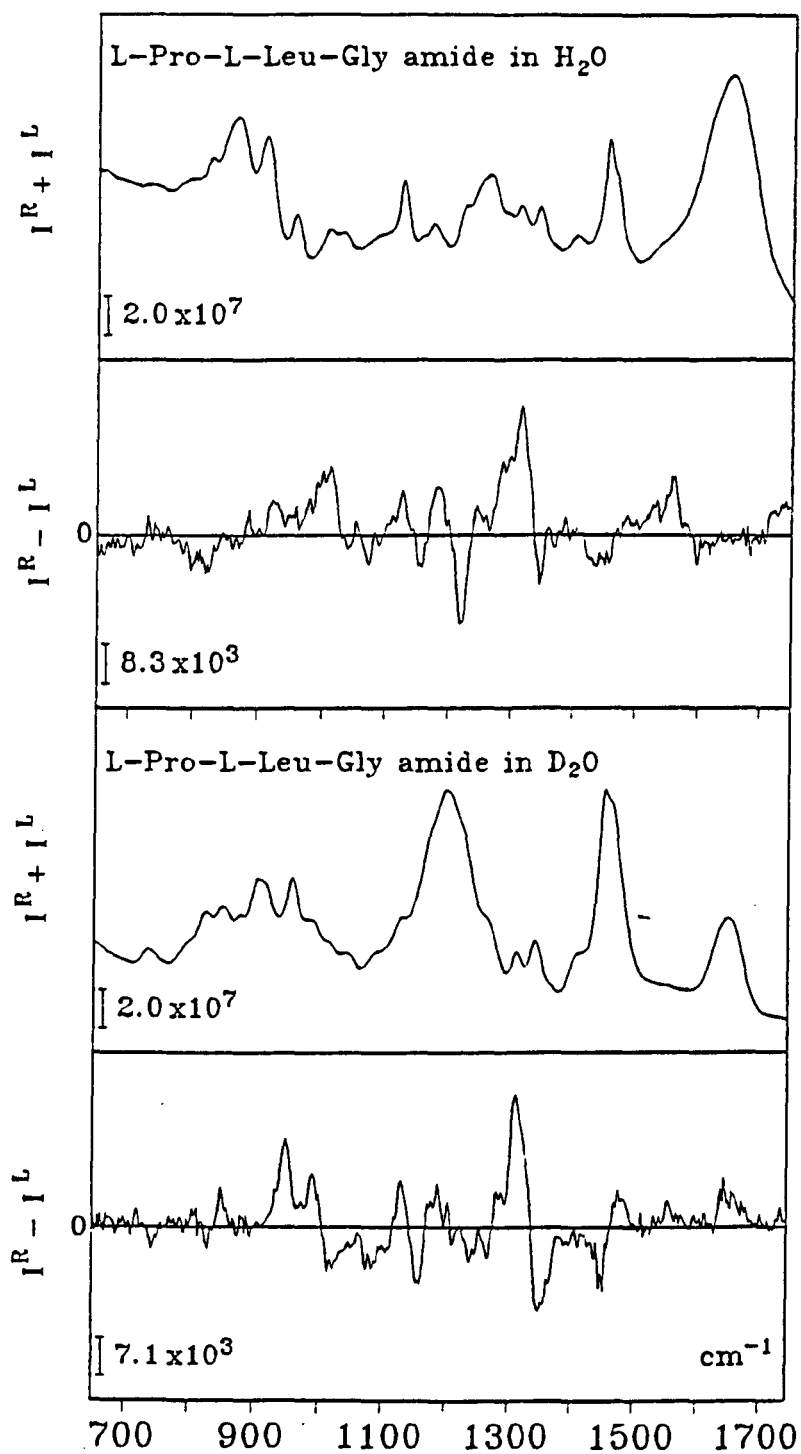


Figure 5.11 Backscattered Raman and ROA of L-Pro-L-Leu-Gly-NH<sub>2</sub>

Basic Raman studies of both the solid and solution state have previously been reported (Hseu & Chang, 1980; Fox *et al.*, 1981) along with a detailed normal mode analysis of the crystalline structure (Naik & Krimm, 1984). Although the conformation of the tripeptide in aqueous solution is not known for certain it is assumed that a rotational isomer is formed that is different to that found in the crystal and dimethylsulphoxide due to hydrogen bonding between the carbonyl groups and surrounding water molecules (Hseu & Chang, 1980).

The ROA features include a negative band at  $1348\text{ cm}^{-1}$ , a strong sharp positive ROA band at  $1317\text{ cm}^{-1}$  and a large negative band at  $1222\text{ cm}^{-1}$ . In  $\text{D}_2\text{O}$  the negative ROA band at  $1348\text{ cm}^{-1}$  undergoes an increase in intensity. As will be discussed in Chapter 6 this is also common in proteins containing  $\beta$ -turns. Assignments made by Naik and Krimm suggest that many of the bands present in the  $\sim 1200 - 1400\text{ cm}^{-1}$  region are due to glycine and leucine  $\text{CH}_2$  wag and twist coordinates, and proline  $\text{CH}_2$  wag coordinates, coupled with in-plane N-H deformations in the crystalline tripeptide. In  $\text{D}_2\text{O}$  this region is very similar since only the amide hydrogens will undergo exchange. Assuming a similar situation in aqueous solution it is reasonable to conclude that negative regions in the  $\sim 1340 - 1380\text{ cm}^{-1}$  region due to  $\text{CH}_2$  deformations coupled with in-plane NH deformations are due to  $\beta$ -turn structures.

## Conclusion

The  $\alpha$ -helical conformation of poly-L-lysine of  $\bar{M}_w \approx 26,000$  provides four major  $\alpha$ -helical ROA signatures: a positively biased couplet in the amide I region centred at  $1645\text{ cm}^{-1}$ ; a "couplet" in the extended amide III region centred at  $1277\text{ cm}^{-1}$ ; a conservative couplet centred at  $1103\text{ cm}^{-1}$  and a broad positive signal between  $\sim 915 - 973\text{ cm}^{-1}$ . As will be discussed later these signals are also present in highly  $\alpha$ -helical proteins. There is also the strong suggestion that the positive ROA band at  $1340\text{ cm}^{-1}$  is due to  $3_{10}$ -helix, which differs from  $\alpha$ -helix since it is more tightly coiled with three residues per turn. This structure is not stable in isolation but is usually found at the ends of  $\alpha$ -helices. The possibility that this feature is due to side-groups can be ruled out, since L-lysine itself exhibits negative ROA in this region. These results suggest that the  $\alpha$ -helical regions in poly-L-lysine are linked by short regions of  $3_{10}$ -helix. Monitoring the  $\alpha$ -helical conformations of poly-L-lysine with  $\bar{M}_w \approx 4,000, 26,000, 56,000$  and  $268,000$  it was found that the amount of  $3_{10}$ -helix increases as the molecular weight increases, suggesting that the  $\alpha$ -helix interconverts with  $3_{10}$ -helix to enable the backbone to negotiate the many more sharp bends expected in the longer polypeptide chain. These observations provide further support for the recently recognised importance of  $3_{10}$ -helix in protein packing (Liu & Day, 1994) and in the local unfolding of proteins.

From the ROA of unordered poly-L-lysine and poly-L-glutamic acid it can be concluded that these so-called "random coil" conformations contain  $\alpha$ -helix and a right-twisted  $\beta$ -strand in a left-handed helix. The two spectra of unordered poly-L-lysine and poly-L-glutamic acid differ sufficiently to assume that they each adopt a different repertoire of structures with the poly-L-glutamic acid appearing to be more disordered. In random coil poly-L-lysine at pH 3, a couplet negative at high wavenumber and positive at low wavenumber centred at  $1108\text{ cm}^{-1}$  and a broad negative band between  $\sim 900 - 971\text{ cm}^{-1}$ , are indicators for the presence of a left-handed helix, since these signals are of the opposite sign to those in a right-handed helix. A left-handed  $\alpha$ -helical backbone is generally not stable over extended regions because the side-chains come into contact with the skeletal backbone. Thus the results appear to support the Tiffany and Krimm model (Tiffany & Krimm, 1968; Krimm & Mark, 1968; Krimm *et al.*, 1969) of short segments of left-handed helix

with approximately 4 - 7 residues in length and 2.5 residues per turn. Random coil poly-L-glutamic acid has a weak negative ROA band between  $\sim 847 - 911 \text{ cm}^{-1}$ , suggesting the presence of left-handed helix but not as much as in the poly-L-lysine. Adding a high concentration of salt and increasing the temperature of the unordered poly-L-lysine the amount of left-handed helix was observed to decrease while the right-handed  $\alpha$ -helix increased.

The model L-Pro-L-Leu-Gly-NH<sub>2</sub> tripeptide provides a useful model to monitor  $\beta$ -turn structure in proteins.

## Chapter 6

### A Survey of Protein ROA Spectra

This chapter contains a discussion of a set of protein ROA spectra that have been measured during the project. These range from proteins that are highly  $\alpha$ -helical to those that have a high  $\beta$ -sheet content. Proteins that contain a mixture of structures will also be discussed. Protein ROA signatures of conformational elements will be proposed. The proteins that will undergo analysis are: lysozyme,  $\alpha$ -lactalbumin, bovine serum albumin (BSA), insulin, ribonuclease A, ovalbumin,  $\beta$ -lactoglobulin, concanavalin A, pepsin,  $\alpha$ -chymotrypsin, trypsin, trypsinogen and orosomucoid. It is useful to run protein ROA spectra in  $D_2O$  as well as in  $H_2O$  because by analysing the exchange of hydrogen atoms on the amide NH groups of the peptide backbone it is possible to assess the relative exposure of different types of structure to the solvent (Pedersen *et al.*, 1991). This is a common procedure in NMR and it is observed that the rate of exchange will decrease rapidly when the NH groups are involved in a hydrogen bonded network. In  $\alpha$ -helix and  $\beta$ -sheet structure the hydrogen bonds therefore restrict exchange as compared to surface loop and turn structure which will exchange rapidly, due to a lack of hydrogen bonding. An ROA spectrum in  $D_2O$  will therefore help to confirm structural predictions and may also provide new information.

The so-called extended amide III region is very important in protein ROA studies because the coupling mechanisms generated between N-H and  $C_\alpha$ -H deformations is particularly sensitive to the geometry (especially the  $\varphi$  angle) and therefore generates a highly informative ROA band structure. The amide II region which occurs in the region  $\sim 1510 - 1570 \text{ cm}^{-1}$  (Diem, 1993) can also provide informative data about  $CH_2$  and  $CH_3$  side-groups and the peptide backbone. Although, amide II bands are not prominent in Raman spectra they are observed in IR and VCD spectra (Gupta & Keiderling, 1992). Protein ROA also shows signals due to side-group vibrations due to for example tryptophan, tyrosine and

phenylalanine.

The material in this chapter is a continuation of the work of Z.Q. Wen (1992; 1994a,b) where tentative assignments were made for  $\alpha$ -helix and  $\beta$ -sheet, and the work of S.J. Ford (1994; 1995a,b) who undertook a study of lysozyme and  $\alpha$ -lactalbumin.

## 6.1 Experimental

Samples of BSA and ribonuclease A were supplied by Fluka; the lysozyme,  $\alpha$ -lactalbumin, insulin, ovalbumin,  $\beta$ -lactoglobulin, concanavalin A, pepsin,  $\alpha$ -chymotrypsin, trypsin, trypsinogen and orosomucoid were obtained from Sigma. All the samples were prepared at ~70 mg/ml in buffered solutions. The lysozyme,  $\alpha$ -lactalbumin, BSA, ovalbumin,  $\beta$ -lactoglobulin and orosomucoid are all in 100 mM acetate buffer at pH 5.4; insulin is in a 100 mM glycine buffer at pH 2.0; ribonuclease A is in a 100 mM acetate buffer at pH 5.0; concanavalin A is in a 100 mM Tris buffer at pH 7.0; pepsin is in a 100 mM Tris buffer at pH 6.0;  $\alpha$ -chymotrypsin is in 100 mM acetate buffer at pH 4.0; trypsin is in a 100 mM Tris buffer at pH 6.0 and the trypsinogen is in a 100 mM Tris buffer at pH 6.6. The D<sub>2</sub>O (purchased from Sigma) samples were made under similar conditions. In all cases the ROA spectra in D<sub>2</sub>O are collected within a few hours after dissolving the solid protein in the required buffer. If the protein is left in D<sub>2</sub>O for a prolonged period the amount of exchange will increase. A detailed deuterium exchange study of lysozyme under different pH's and time periods has previously been performed (Ford, 1995a).

All samples were treated with a small amount of pharmaceutical grade charcoal to remove fluorescing impurities and then filtered into quartz microfluorescence cells through 0.22  $\mu$ m Millipore membrane filters. Laser excitation at 514.5 nm was used, with power of ~700 mW at the sample. Each ROA spectra was acquired over ~10 h.



## 6.2 Results and Discussion

Each of the protein ROA spectra will now be discussed and where possible comparisons will be made with structures that are known to be similar from other biophysical techniques.

### Lysozyme and $\alpha$ -Lactalbumin

Hen egg-white lysozyme and bovine  $\alpha$ -lactalbumin are ideal samples for initial detailed ROA studies of proteins due to the wealth of information available from X-ray, NMR and other biophysical techniques. Due to previous comparisons of these two proteins (McKenzie & White, 1991) they will be directly compared in this subsection.

Hen egg-white lysozyme is an anti-bacterial enzyme with a  $\beta(1 - 4)$  glucoaminidase activity, allowing it to catalyse hydrolysis the walls of bacterial cells. Hen egg-white lysozyme is a member of the c-type lysozyme family which includes all of the mammalian lysozymes, moth lysozymes and some bird lysozymes.

Hen egg-white lysozyme contains 129 amino acid residues and has a molecular weight of 14.3 kD. The structure determined by X-ray crystallography (Blake *et al.*, 1965; Blake *et al.*, 1981) is illustrated in Figure 6.1, drawn using the program Molscript (Kraulis, 1991).

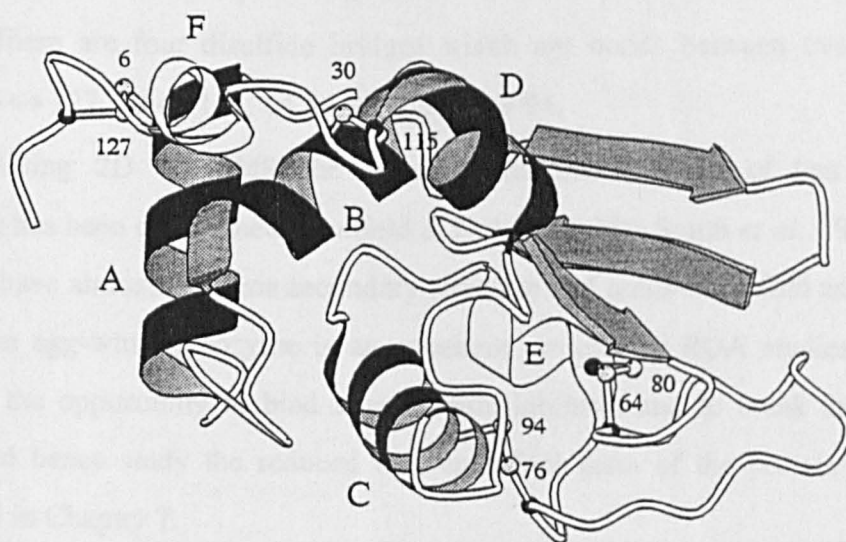


Figure 6.1 Molscript drawing of lysozyme

The protein itself is made up of an  $\alpha$ - and a  $\beta$ -domain. The actual secondary structure composition is listed in Table 6.1.

Table 6.1 Secondary structure of hen egg-white lysozyme  
(McKenzie & White, 1991)

Structure	Residues
$\alpha$ -helix	4 - 15(A), 24 - 36(B), 88 - 99(C), 108 - 115 (D)
$3_{10}$ -helix	79 - 84 (E), 119 - 124 (F)
$\beta$ -sheet	1 - 2, 39 - 40 (double-stranded) 42 - 46, 50 - 54, 57 - 60 (triple-stranded)

As can be seen from Table 6.1 and Figure 6.1 there are four  $\alpha$ -helices which encompass both the amino- and carboxy-terminal segments, making up the  $\alpha$ -domain of the lysozyme. The  $\beta$ -domain is made up of a triple-stranded antiparallel  $\beta$ -sheet, a short double-stranded antiparallel  $\beta$ -sheet that links the two domains and a loop region. There are four disulfide bridges which are bonds between two cysteines, namely  $6 \leftrightarrow 127$ ,  $30 \leftrightarrow 115$ ,  $64 \leftrightarrow 80$  and  $76 \leftrightarrow 94$ .

Using 2D  $^1\text{H}$  NMR the aqueous solution structure of hen egg-white lysozyme has been determined (Redfield & Dobson, 1988; Smith *et al.*, 1993). It was found to have almost the same secondary structure and main-chain fold as the crystal form. Hen egg-white lysozyme is an excellent protein for ROA studies as it also provides the opportunity to bind a saccharide inhibitor and to break the disulfide bonds and hence study the reduced (i.e. unfolded) form of the protein as will be discussed in Chapter 7.

The other protein to be discussed in this subsection is  $\alpha$ -lactalbumin which

is present in the milk of most mammals where it combines with galactotransferase to switch its catalytic function from the production of N-acetyllactosamine to the production of lactose.

Bovine  $\alpha$ -lactalbumin which is the form examined here contains 123 amino acid residues and has a molecular weight of 14.2 kDa. Both the structures of baboon and human have been determined by X-ray crystallography (Acharya *et al.*, 1989, 1991). It is unfortunate that the X-ray structure of the bovine form is not known but NMR studies (Alexandrescu *et al.*, 1992) suggest that the solution structure is similar to that of the human form, which is also similar to that of the baboon  $\alpha$ -lactalbumin (Acharya *et al.*, 1991). In Figure 6.2 the structure of baboon  $\alpha$ -lactalbumin is shown.

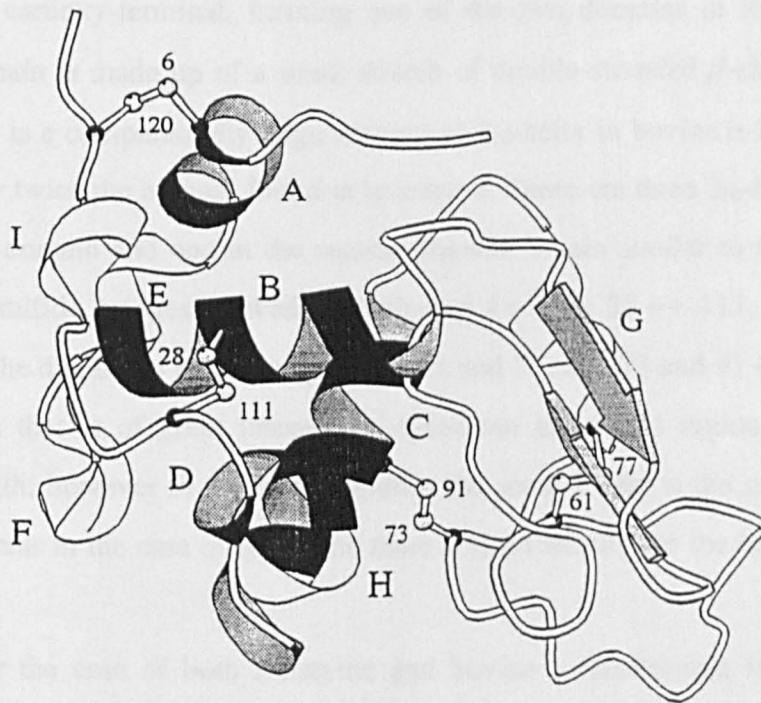


Figure 6.2 Molscript drawing of  $\alpha$ -lactalbumin

The X-ray crystal structures of lysozyme and  $\alpha$ -lactalbumin are very similar, attracting comparisons of the structure and function of both proteins (McKenzie & White, 1991; Sugai & Ikeguchi, 1994). The arrangement of the secondary structure elements of  $\alpha$ -lactalbumin are listed in Table 6.2.

Table 6.2 Secondary structure of  $\alpha$ -lactalbumin (McKenzie & White, 1991)

Structure	Residues
$\alpha$ -helix	5 - 11 (A), 23 - 34 (B), 86 - 99 (C), 105 - 109 (D)
$3_{10}$ -helix	12 - 16 (E), 17 - 21 (F), 76 - 82 (G), 101 - 104 (H), 115 - 119 (I)
$\beta$ -sheet	40 - 43, 47 - 50 (double-stranded)

Similar to lysozyme there are four  $\alpha$ -helix regions which encompass the amino and carboxy-terminal, forming one of the two domains in the protein. The second domain is made up of a small stretch of double-stranded  $\beta$ -sheet and a long loop. There is a comparatively large amount of  $3_{10}$ -helix in bovine  $\alpha$ -lactalbumin; in fact roughly twice the amount found in lysozyme. There are three  $3_{10}$ -helix regions in the  $\alpha$ -helix domain and one in the second domain. Again similar to lysozyme there are four disulfide bridges between cysteines 6  $\leftrightarrow$  120, 28  $\leftrightarrow$  111, 61  $\leftrightarrow$  77 and 73  $\leftrightarrow$  91. The disulfides between cysteines 61 and 77, and 73 and 91 create a double loop region that is of great interest. In lysozyme there is a region of equivalent overall length, however in the  $\alpha$ -lactalbumin the second loop is the calcium binding region whereas in the case of lysozyme there is little affinity for the binding of metal ions.

For the case of both lysozyme and bovine  $\alpha$ -lactalbumin it is also worth noting that early conventional Raman studies (Yu, 1974) showed that the crystal and solution main-chain conformations of both proteins were very similar.

The backscattered Raman and ROA spectra of  $\alpha$ -lactalbumin and lysozyme in  $\text{H}_2\text{O}$  and  $\text{D}_2\text{O}$  and are shown in Figure 6.3 and Figure 6.4, respectively.

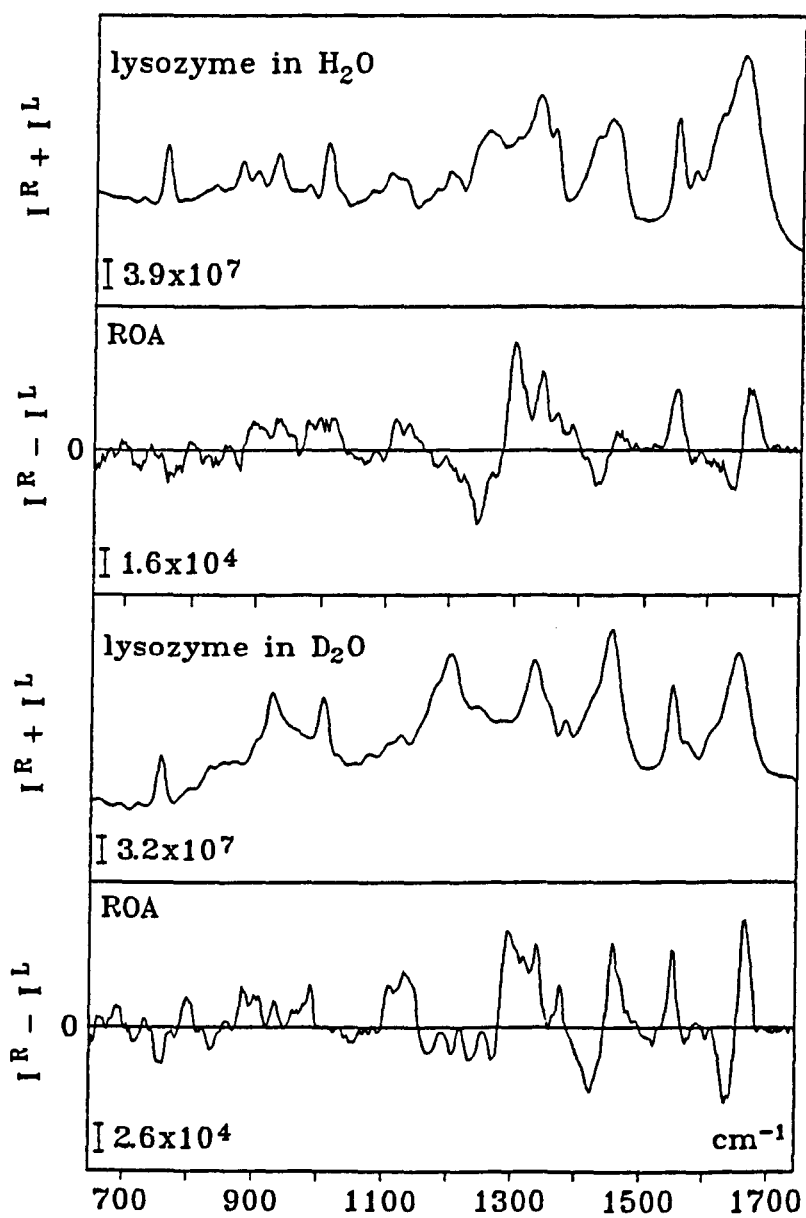


Figure 6.4 Backscattered Raman and ROA of lysozyme

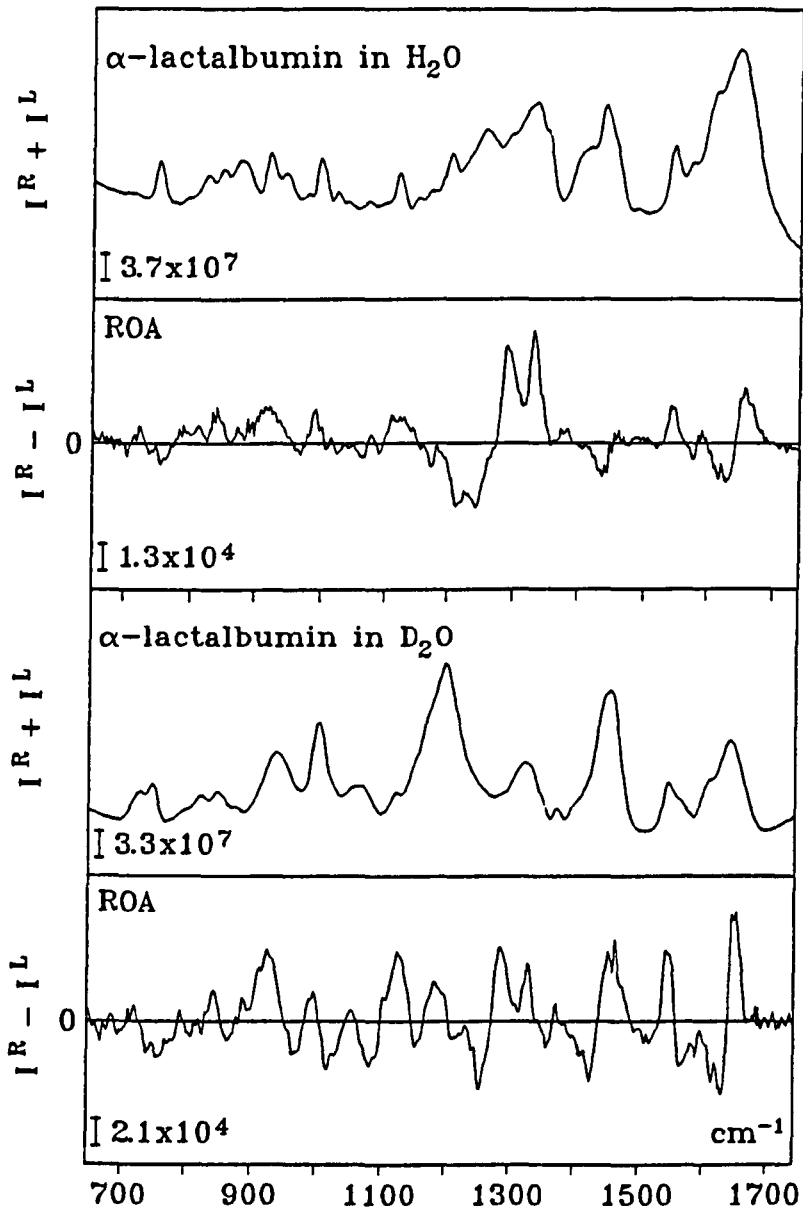


Figure 6.5 Backscattered Raman and ROA of  $\alpha$ -lactalbumin

The backbone skeletal stretch region of lysozyme and  $\alpha$ -lactalbumin contain valuable information about the  $\alpha$ -helix (Wen *et al.*, 1994a) and  $\beta$ -sheet (Wen *et al.*, 1994b) regions of the proteins. Both  $\alpha$ -lactalbumin and lysozyme contain very similar ROA bands in this region since both contain a positive region in the range  $\sim 880 - 960 \text{ cm}^{-1}$ , which is characteristic of  $\alpha$ -helix modes. There are nonetheless differences since lysozyme contains two distinct ROA peaks at  $910$  and  $940 \text{ cm}^{-1}$ , while  $\alpha$ -lactalbumin contains a positive ROA band at  $855 \text{ cm}^{-1}$ . Also  $\alpha$ -lactalbumin contains a positive band at  $1006 \text{ cm}^{-1}$  which is probably due to twisted  $\beta$ -sheet (see later) or possibly aromatic side-groups. Lysozyme contains a similar positive band at  $1020 \text{ cm}^{-1}$ , again probably originating in twisted  $\beta$ -sheet. A second  $\alpha$ -helix ROA band structure which is observed in both proteins is a couplet which has a small negative component at  $1095 \text{ cm}^{-1}$  with a large positive component at  $1120 \text{ cm}^{-1}$ . In  $\text{D}_2\text{O}$  the  $\alpha$ -lactalbumin ROA spectrum has a large positive region in the range  $\sim 882 - 949 \text{ cm}^{-1}$  and a large negative-positive couplet with the negative part peaking at  $1084 \text{ cm}^{-1}$  and the positive part at  $1124 \text{ cm}^{-1}$ , indicating that the  $\alpha$ -helix has undergone little exchange. The  $\beta$ -sheet region is quite different in  $\text{D}_2\text{O}$ . There is a similar situation in the lysozyme  $\text{D}_2\text{O}$  ROA spectrum with a positive ROA region between  $\sim 875 - 960 \text{ cm}^{-1}$  and the  $\alpha$ -helix negative-positive couplet with a crossover point at  $1097 \text{ cm}^{-1}$  still being observable. It can thus be concluded from the overall fingerprints of ROA in this region that  $\alpha$ -lactalbumin and lysozyme contain similar secondary structure in aqueous solution. The small differences that do occur could be significant since they might be due to for example the different lengths of the  $\alpha$ -helical segments and their precise environments in the protein. The larger amount of  $\beta$ -sheet in lysozyme will also affect the ROA.

The ROA of  $\alpha$ -lactalbumin in  $\text{H}_2\text{O}$  in the range  $\sim 950 - 1070 \text{ cm}^{-1}$  contains a negative-positive-negative ROA pattern which may be indicating that there are residues in solution that have the propensity to form a right-twisted  $\beta$ -strand in a left-handed helical conformation. Lysozyme also has hints of this pattern.

ROA is therefore suggesting that  $\alpha$ -lactalbumin and lysozyme have very similar secondary structure in aqueous solution. It is interesting however that this

result is in conflict with a recent VCD result which suggested that the secondary structures of  $\alpha$ -lactalbumin and lysozyme in aqueous solution are much less similar than in the crystal (Keiderling *et al.*, 1991).

Unfortunately, ROA is not particularly useful in the amide I region. It is found that both  $\alpha$ -helix and  $\beta$ -sheet give rise to an amide I ROA couplet, negative at low wavenumber and positive at high. In  $\beta$ -sheet proteins this couplet tends to be conservative while in  $\alpha$ -helix proteins the positive component has more intensity. This situation is in contrast to VCD where the amide I region is very sensitive to different types of protein secondary structure (Keiderling & Pancoska, 1993; Keiderling *et al.*, 1994; Pancoska *et al.*, 1994). Comparing the amide I ROA region of lysozyme and  $\alpha$ -lactalbumin, the amide I ROA couplets can be seen to be very similar apart from the  $\alpha$ -lactalbumin having an extra smaller couplet at lower frequency (due to aromatic aromatic side-groups). This observation again highlights the similarity of the secondary structure in both proteins. Although the amide I region also contains information about loops and turns (Ford *et al.*, 1995b), deconvolution of this band would be required to fully extract this data. Due to secondary structure contributing most to the amide I band, in D<sub>2</sub>O this band is very similar to the H<sub>2</sub>O amide I region, although in both  $\alpha$ -lactalbumin and lysozyme the ROA couplet appears to be slightly sharper (probably due to exposed loop structure being exchanged).

In the extended amide III region protein ROA has started to reveal detailed information of not only the secondary but also the tertiary structure of proteins. Although most proteins have a broad couplet in this region, negative at low wavenumber and positive at high, the actual detailed structure illustrates the incisive nature of ROA. Both lysozyme and  $\alpha$ -lactalbumin have a strong sharp positive ROA signal at 1296 cm<sup>-1</sup> with a negative shoulder at 1265 cm<sup>-1</sup>, which has previously been assigned to  $\alpha$ -helix (Wen *et al.*, 1994a). In both proteins this band is relatively unaffected in D<sub>2</sub>O, providing further indication that there has been little exchange of the  $\alpha$ -helix. As will be discussed later  $\beta$ -sheet proteins show a sharp positive band at 1313 cm<sup>-1</sup> and it is noticeable that this band is absent in both lysozyme and  $\alpha$ -



lactalbumin, probably due to the low  $\beta$ -sheet content in both proteins. There is also the possibility that the  $1313\text{ cm}^{-1}$   $\beta$ -sheet band is being wiped out due to the other intense bands around it. The detailed structure in the extended amide III region also allows ROA to probe the differences in the aqueous solution structure of  $\alpha$ -lactalbumin and lysozyme. The sharp positive band at  $1340\text{ cm}^{-1}$  observed in both  $\alpha$ -lactalbumin and lysozyme which was also observed in the  $\alpha$ -helical poly-L-lysine spectra has been attributed to  $3_{10}$ -helix. Evidence for this assignment comes from the fact that X-ray studies have shown  $\alpha$ -lactalbumin to have twice the amount of  $3_{10}$ -helix as compared to lysozyme and indeed the ROA band is roughly twice as large in  $\alpha$ -lactalbumin. In  $\text{D}_2\text{O}$  although the  $1340\text{ cm}^{-1}$  ROA band has undergone deuterium exchange the band is still observable, due to some of the  $3_{10}$ -helix being buried in the structure. In the negative half of the couplet  $\alpha$ -lactalbumin has two sharp negative ROA bands at  $1223$  and  $1243\text{ cm}^{-1}$ , while lysozyme has one sharp band at  $1237\text{ cm}^{-1}$  (a possible assignment for this band will be proposed in Chapter 7). In  $\text{D}_2\text{O}$  both of these ROA bands have undergone almost complete exchange highlighting that this part of the protein structure is exposed to the solvent. These differences indicate that the details of the loop structure of the two proteins are different.

From the ROA study of the unfolded lysozyme which will be discussed later in Chapter 7, the underlying amide III couplet which can be observed in both  $\alpha$ -lactalbumin and lysozyme is probably due to the loop and end-chain structure that has the same conformational heterogeneity and associated flexibility as that of an unfolded protein.

Protein ROA can also provide valuable information about side-group conformation and this will now be discussed for  $\alpha$ -lactalbumin and lysozyme.  $\alpha$ -lactalbumin contains four tryptophan residues while lysozyme contains six tryptophan residues. Both proteins contain a strong positive ROA signal associated with the strong Raman band at  $1555\text{ cm}^{-1}$  and weak negative ROA associated with the weaker band at  $1580\text{ cm}^{-1}$ . As expected the protein ROA bands are relatively unaffected in  $\text{D}_2\text{O}$ , due to the side-groups not exchanging.  $\alpha$ -lactalbumin and lysozyme also show a strong Raman band at  $1451\text{ cm}^{-1}$  accompanied by a shoulder at

1420  $\text{cm}^{-1}$  in  $\alpha$ -lactalbumin and at 1430  $\text{cm}^{-1}$  in lysozyme. These bands mostly originate in side-group  $\text{CH}_2$  and  $\text{CH}_3$  deformations and also a small contribution from the N-H bend of the indole ring in the tryptophan to the 1420 and 1430  $\text{cm}^{-1}$  shoulders (Harada & Takeuchi, 1986). The ROA of both proteins in this region consists of an ROA couplet that is large and negative at the low wavenumber region. The positive parts of the ROA couplet have no significant Raman bands associated with them but it is known that vibrations of the amide II peptide backbone and of some aromatic residues occur in this region because they have been observed to become very strong in surface-enhanced Raman spectra (Hu *et al.*, 1995). A possible explanation is that the ROA band pattern arises from the coupling of side-group  $\text{CH}_2$  and  $\text{CH}_3$  deformations with the previously mentioned backbone and side-group modes. Due to the side-groups not exchanging these bands are again unaffected in  $\text{D}_2\text{O}$ . One difference in the ROA of  $\alpha$ -lactalbumin and lysozyme is that  $\alpha$ -lactalbumin has a small sharp couplet, positive at low wavenumber and negative at high which has a crossover point of 1615  $\text{cm}^{-1}$ . This ROA feature is on the low-wavenumber side of the amide I Raman band that has previously been assigned to aromatic vibrations such as tryptophan (Lord & Yu, 1970; Harada & Takeuchi, 1986). The ROA data is therefore showing that in aqueous solution that there are differences in the conformations of some of the aromatic and aliphatic side-groups.

### Bovine Serum Albumin (BSA)

Serum albumin is a principal component of blood and contributes about 80% to the colloid osmotic blood pressure. It is also responsible for the maintenance of blood pH (Figge *et al.*, 1991). Although this study will concentrate on the bovine form it is only the human form that has had its structure solved by crystallography (He & Carter, 1992).

BSA is a large protein consisting of 582 amino acids (Brown, 1975; Holowachuk, 1991) with a molecular weight of ~65 kDa containing 17 disulfide bridges. Although the bovine crystal structure is not known, it is assumed that it is

similar to the human form. Unfortunately, the atomic coordinates of human serum albumin (HSA) have not been released due to the structure having pharmaceutical applications and therefore a Molscript diagram cannot be used to illustrate the structure. Instead a topological illustration of a typical domain in HSA is shown in Figure 6.5. The protein is constructed from three homologous domains that assemble to form a heart-shaped molecule.

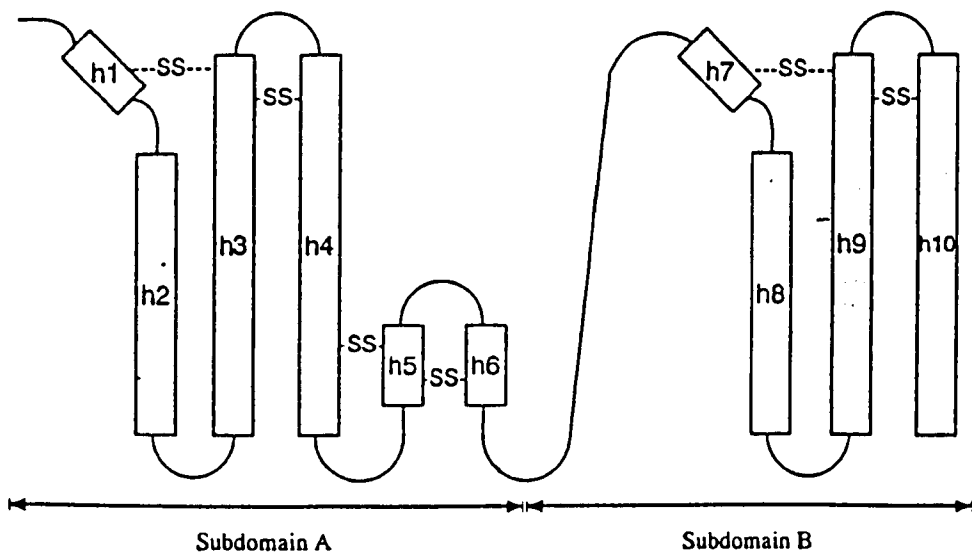


Figure 6.5 Topological illustration of a typical domain in HSA (He & Carter, 1992)

The structure is highly helical, containing 67% helix. In Figure 6.5 there are 10 principal helices labelled h1 to h10. Three domains which are similar to that in Figure 6.5 make up the whole of the protein and it is therefore made up of six helical subdomains (Carter *et al.*, 1989; Carter & He, 1990). The helices (i.e. h1, h2, h3, h4, h7, h8, h9 and h10) represent the common structural motif that is common in all six subdomains. The motif is therefore a short helix (e.g. h1) connected to three antiparallel long helices (e.g. h2, h3 and h4) that are held together by a pair of disulfide bridges.

The ROA of BSA will now be discussed. In Figure 6.6 there are the backscattered Raman and ROA spectra of BSA in H<sub>2</sub>O and D<sub>2</sub>O at pH/pD 5.4. Comparing the ROA spectrum of BSA in H<sub>2</sub>O with that of the  $\alpha$ -helical poly-L-lysine in H<sub>2</sub>O previously discussed, there are three  $\alpha$ -helical signatures that can be clearly observed, namely the positive ROA region of  $\sim 851 - 957 \text{ cm}^{-1}$ , the negative-positive ROA couplet with a crossover at  $1103 \text{ cm}^{-1}$  and the negative-positive underlying ROA couplet in the amide III region with a crossover at  $1277 \text{ cm}^{-1}$  (Wen *et al.*, 1994a). It can also be deduced from the  $\alpha$ -helical poly-L-lysine that the positively biased negative-positive amide I couplet with a crossover at  $1641 \text{ cm}^{-1}$  is also an  $\alpha$ -helical signature. It is reassuring that in D<sub>2</sub>O these four features are relatively unchanged providing confidence in the prediction that they are indeed signatures for  $\alpha$ -helix.

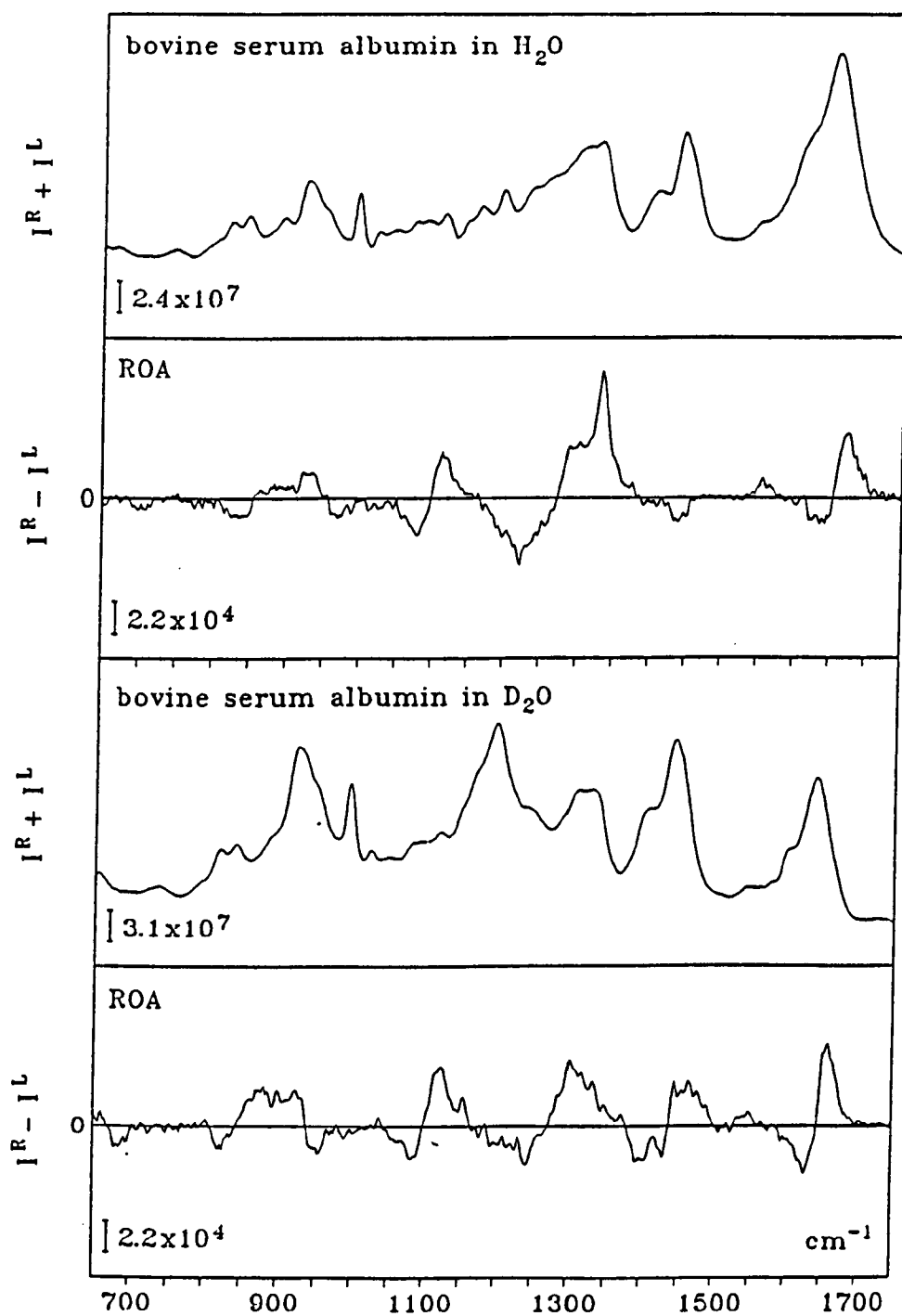


Figure 6.6 Backscattered Raman and ROA of BSA

The strong positive ROA band at  $1340\text{ cm}^{-1}$  which is also found in the  $\alpha$ -helical poly-L-lysine has been postulated as  $3_{10}$ -helix. It is obvious that in  $\text{D}_2\text{O}$  this band has undergone almost complete exchange, providing confidence in the assignment because  $3_{10}$ -helix is known to be prone to exchange due to the lack of strong hydrogen bonding. There is also the loss of the negative ROA band at  $1223\text{ cm}^{-1}$  in  $\text{D}_2\text{O}$ . It is interesting that native  $\alpha$ -lactalbumin also has a negative ROA band at this precise frequency. There must therefore be some sort of common loop structure present in  $\alpha$ -lactalbumin and BSA (but absent in lysozyme).

The small positive ROA band at  $1551\text{ cm}^{-1}$  is due to the two tryptophans in the protein. In  $\text{D}_2\text{O}$  this band shifts down slightly and becomes even broader.

In the  $\text{H}_2\text{O}$  BSA spectrum there is a small negative ROA band between  $\sim 837 - 881\text{ cm}^{-1}$  and a small negative ROA band between  $\sim 1418 - 1456\text{ cm}^{-1}$  which is due to  $\text{CH}_2$  and  $\text{CH}_3$  side-groups. Other spectroscopic methods (Jacobsen, 1972; Sjöholm & Ljunstedt, 1973; Chen & Lord, 1976) also suggested the solution structure to be predominantly  $\alpha$ -helical.

## Insulin

Insulin is a central component in the hormonal control of metabolism (Cahill, 1971) and there is much interest in the structure and function of the protein (Blundell & Humbel, 1980). Insulin is an ideal subject for an ROA study because the crystal structure has been solved (Chothia *et al.*, 1983; Baker *et al.*, 1988; Derewenda *et al.*, 1989; Wlodawer & Savage, 1989; Bentley *et al.*, 1991) and there has been extensive attempts to determine the structure and dynamics of insulin in solution (Hua *et al.*, 1991). Insulin has also undergone Raman spectroscopic studies (Yu *et al.*, 1972; Yu & Liu, 1972).

The structure of insulin is illustrated by a Molscrip diagram in Figure 6.7.

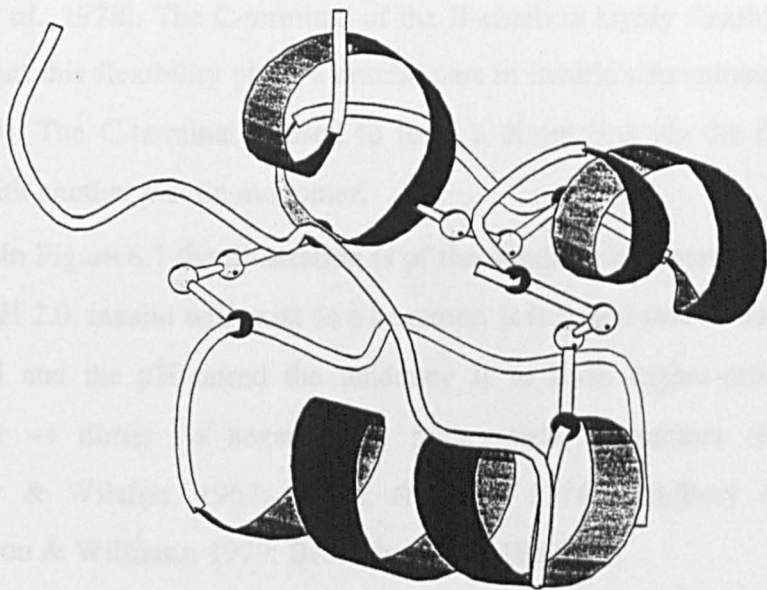


Figure 6.7 Molscript drawing of insulin

Insulin is made up of two polypeptide chains: an A-chain with 21 residues and a B-chain with 30 residues. These two polypeptide chains are linked by two disulfide bonds (A7 - B7 and A20 - B19) and there is a third intramolecular disulfide bond in the A-chain (A6 - A11). Table 6.3 summarizes the structure of the protein.

Table 6.3 Structural elements in the monomer of insulin. (Hua *et al.*, 1991).

Structure	Residues
Helix	A2 - A8 (H1), A12 - A18 (H2), B30 - B40 (H3)
Turn	B20 - B23 (T1)

The C-terminal on the B-chain which includes residues B23 - B26 has been found to play an important role in the function of insulin (Pullen *et al.*, 1976; De Meyts *et al.*, 1978). The C-terminal of the B-chain is highly flexible and it has been shown that this flexibility plays a crucial part in insulin's functional role (Dodson *et al.*, 1979). The C-terminal is used to form a dimer link *via* the formation of a  $\beta$ -strand with another insulin monomer.

In Figure 6.7 the illustration is of the insulin monomer; however at  $\sim 70$  mg/ml and pH 2.0, insulin will exist as a hexamer. It is found that as the concentration is increased and the pH raised the tendency is to form higher-order oligomers i.e. monomer  $\rightarrow$  dimer  $\rightarrow$  hexamer  $\rightarrow$  higher-order oligomers (Kowalsky, 1962; Bradbury & Wilairat, 1967; Paselk & Levy, 1974; Bradbury & Brown, 1977; Williamson & Williams, 1979; Bradbury *et al.*, 1981).

The backscattered Raman and ROA spectra of insulin in  $\text{H}_2\text{O}$  and  $\text{D}_2\text{O}$  are shown in Figure 6.8. From the ROA it is clear that insulin is a highly  $\alpha$ -helical protein. This agrees with crystallography where it is known that insulin exists in a compact globular structure with an  $\alpha$ -helix in the centre of the B-chain (B10 - B19) and two shorter helices in the A-chain (A1 - A6 and A16 - A21) which are linked by a small extended chain (A7 - A13). Thus in the  $\text{H}_2\text{O}$  spectrum the positive ROA band at  $\sim 860 - 974 \text{ cm}^{-1}$ , negative-positive ROA couplet with a crossover at  $1101 \text{ cm}^{-1}$ , and the underlying negative-positive ROA couplet in the amide III region with a crossover at  $1277 \text{ cm}^{-1}$  all indicate that the protein is composed mainly from  $\alpha$ -helix. The amide I negative-positive ROA couplet with the negative part peaking at  $1638 \text{ cm}^{-1}$  and the positive at  $1668 \text{ cm}^{-1}$ , is only slightly positive biased and is unfortunately not very informative. In  $\text{D}_2\text{O}$  the three  $\alpha$ -helical signatures are virtually unchanged, as would be expected for secondary structure.

In the positive part of the amide III there is a hint of a shoulder at  $1340 \text{ cm}^{-1}$ , perhaps suggesting that there is a small amount of  $3_{10}$ -helix present in the solution structure. As expected this disappears in the  $\text{D}_2\text{O}$  spectrum.



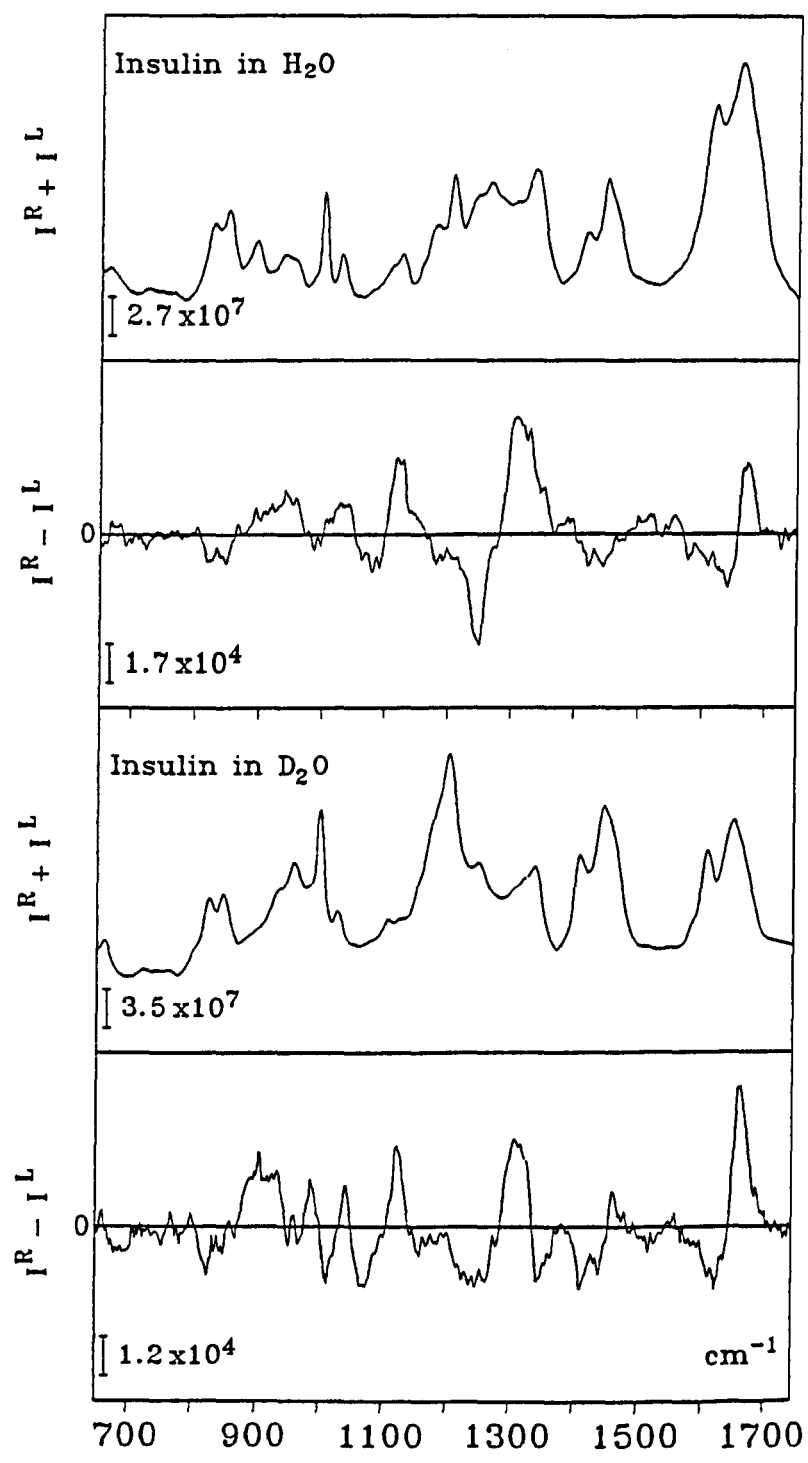


Figure 6.8 Backscattered Raman and ROA of insulin

The sharp negative ROA band in the H<sub>2</sub>O spectrum at 1243 cm<sup>-1</sup> which undergoes exchange in D<sub>2</sub>O is indicative of a particular type of loop structure that is also present in  $\alpha$ -lactalbumin.

It is also noticeable that there is no tryptophan band in the ROA spectrum, confirming this assignment because there is indeed no tryptophan present in insulin.

### Ribonuclease A

Ribonuclease A is a digestive enzyme secreted by the pancreas which catalyzes the hydrolysis of phosphodiester bonds in ribonucleic acid (i.e. RNA). The molecular weight of the protein is ~13.6 kDa, containing 124 residues.

Ribonuclease A is an excellent sample for ROA studies because not only is the X-ray crystal structure known (Avey *et al.*, 1967; Kartha *et al.*, 1967; Borkakoti *et al.*, 1982; Wlodawer & Sjölin, 1983; Svensson *et al.*, 1986; Wlodawer *et al.*, 1986, 1988; Campbell & Petsko, 1987) but the solution structure has also been solved using 2-dimensional NMR (Santoro *et al.*, 1993).

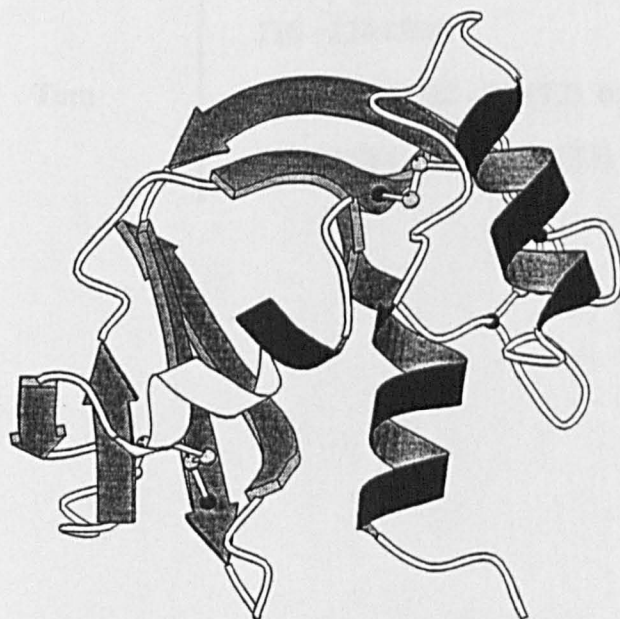


Figure 6.9 Molscript drawing of ribonuclease A

As can be observed from Figure 6.9, ribonuclease A contains a mixture of both helix and sheet. The structure consists of three main helical sections:  $\alpha$ -helix from residue 4 - 13;  $\alpha$ -helix from residue 24 - 33 and a helical section from residue 51 - 59 which is irregular, beginning with one turn of  $\alpha$ -helix followed by two turns of  $3_{10}$ -helix. There are two main  $\beta$ -sheet regions which both have their strands in an antiparallel arrangement that is twisted. There is also a considerable amount of turn structure present (Santoro *et al.*, 1993). Table 6.4 lists the main structural features in ribonuclease A.

Table 6.4            Structural elements present in solution ribonuclease A  
(Santoro *et al.*, 1993)

Structure	Residues
$\alpha$ -helix	4 - 13 (H1), 24 - 33 (H2), 51 - 54 (H3)
$3_{10}$ -helix	54 - 59 (H3)
$\beta$ -sheet	42 - 48 (S11), 79 - 87 (S12), 96 - 104 (S13), 61 - 64 (S21), 71 - 74 (S22), 105 - 111 (S23), 116 - 124 (S24)
Turn	14 - 17 (T1), 22 - 25 (T2), 65 - 68 (T3), 91 - 94 (T4), 112 - 115 (T5)

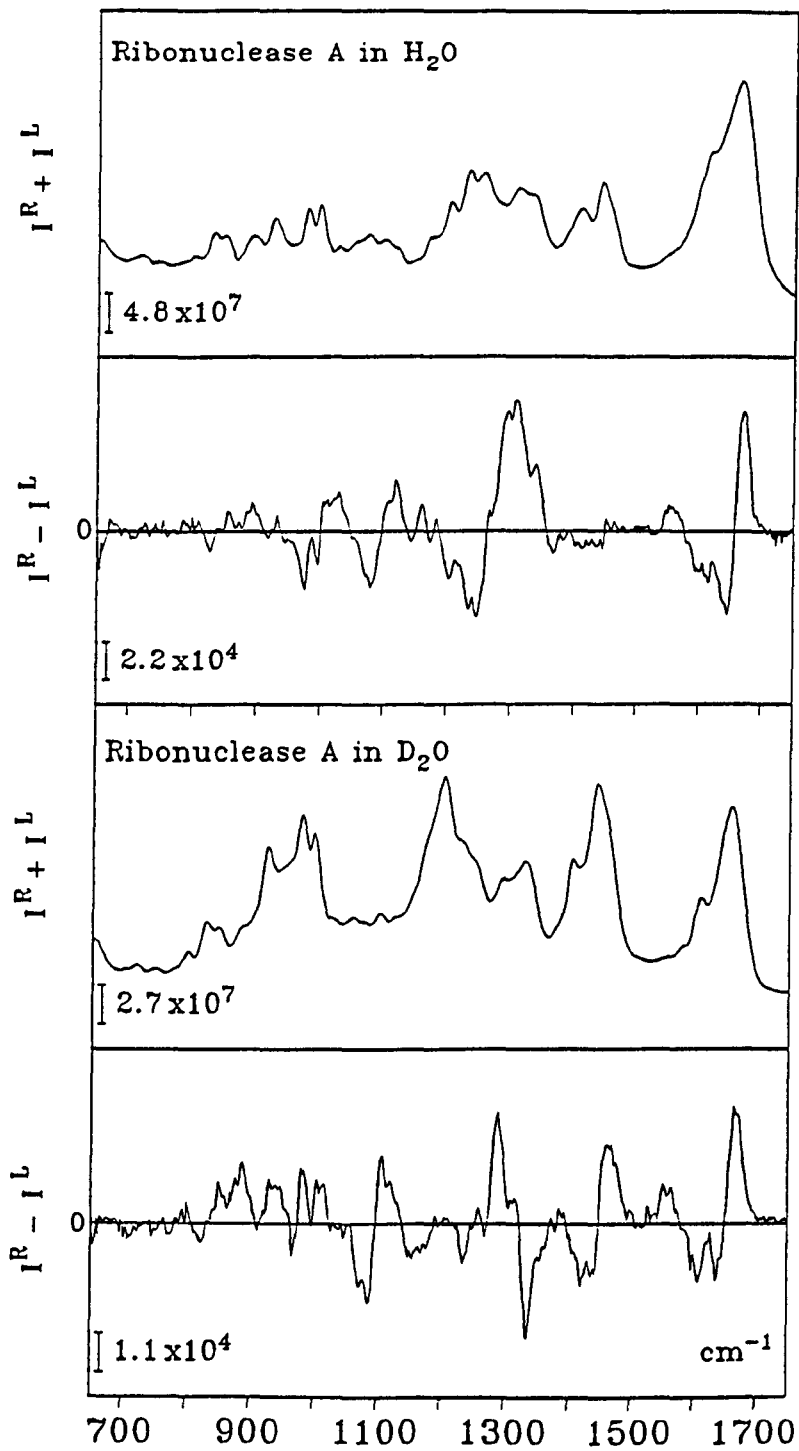


Figure 6.10 Backscattered Raman and ROA of ribonuclease A

The backscattered Raman and ROA spectra of ribonuclease A in H<sub>2</sub>O and D<sub>2</sub>O are shown in Figure 6.10. There are again indicators of  $\alpha$ -helix being present: the positive ROA region between  $\sim 831 - 909 \text{ cm}^{-1}$ , the negative-positive couplet with a crossover at  $1101 \text{ cm}^{-1}$  and a positively biased amide I couplet with the negative part peaking at  $1649 \text{ cm}^{-1}$  and the positive part peaking at  $1673 \text{ cm}^{-1}$ . In D<sub>2</sub>O these three ROA bands are all virtually unchanged. The negative-positive couplet with a crossover point at  $1267 \text{ cm}^{-1}$  is also an  $\alpha$ -helix signature which is again only slightly altered in the D<sub>2</sub>O spectrum. There is also a sharp peak at  $1340 \text{ cm}^{-1}$  indicating the presence of  $3_{10}$ -helix agreeing with the NMR solution structure (Santoro *et al.*, 1993), which loses most of its intensity in D<sub>2</sub>O.

Comparing ribonuclease A to other proteins with a high content of  $\beta$ -sheet (see later), common ROA signatures are again observed. The positive ROA region between  $\sim 1001 - 1048 \text{ cm}^{-1}$  and the sharp positive ROA peak at  $1311 \text{ cm}^{-1}$  are all indicators of  $\beta$ -sheet (Wen *et al.*, 1994b). The negative-positive-negative pattern in the range of  $\sim 950 - 1070 \text{ cm}^{-1}$  as suggested before may not only be indicating the presence of  $\beta$ -sheet but that the sheets are formed from right-twisted  $\beta$ -strands in a left-handed helical conformation as can be observed in the Molscript illustration of the protein. The  $\beta$ -sheets in ribonuclease A are allowed to twist since they are highly irregular with only up to four strands of different lengths in each of the sheets. This is in contrast to concanavalin A, where the sheets remain flat in a multi-stranded  $\beta$ -sheet structure forming a  $\beta$ -barrel (see later). In D<sub>2</sub>O there is much more exchange of the  $\beta$ -sheet structure than the  $\alpha$ -helix structure. The positive ROA band in the H<sub>2</sub>O spectrum between  $\sim 1001 - 1048 \text{ cm}^{-1}$  has split into two sharp peaks in the D<sub>2</sub>O spectrum. It is also noticeable that the  $1311 \text{ cm}^{-1}$   $\beta$ -sheet band has lost over half of its intensity in D<sub>2</sub>O. It must therefore be remembered that although the amide hydrogens that are internally hydrogen bonded in the  $\beta$ -sheet network are protected from exchange, the hydrogens exposed on the outside of this network are readily exchanged.

In D<sub>2</sub>O there is a sharp negative ROA band at  $1337 \text{ cm}^{-1}$  which although at a slightly different frequency from that in the model  $\beta$ -turn peptide, could be indicating

the presence of turn structure. The advantage of being able to run D<sub>2</sub>O spectra is therefore highlighted here since the band is 'masked' in the H<sub>2</sub>O spectrum.

### Ovalbumin

Ovalbumin is the major protein in avian egg-white and in fact comprises 60 - 65% of the total protein in egg-white although its function is unknown. Ovalbumin is a glycoprotein with a small amount of carbohydrate and has a molecular mass of ~45 kDa, comprising 385 residues (McReynolds *et al.*, 1978).

The protein exists as a tetramer but due to the size of the final structure only the monomer is shown in Figure 6.11. Although the four molecules making up the final ovalbumin model are all slightly different the differences are not significant (Millar *et al.*, 1983; Stein *et al.*, 1991). A summary of the structure of the monomer is given in Table 6.5.

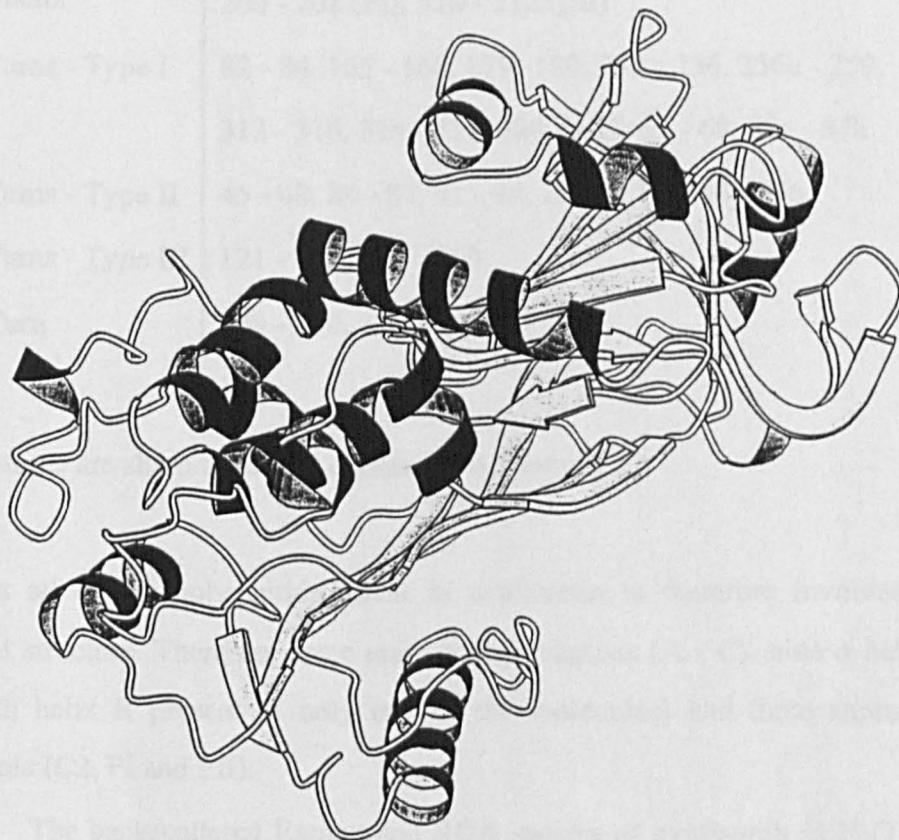


Figure 6.11 Molscript drawing of ovalbumin

Table 6.5      Structural summary of one of the four molecules making up ovalbumin (Stein *et al.*, 1991).

Structure	Residues
$\alpha$ -helix	26 - 44 (A), 54 - 65 (B), 70 - 80 (C), 87b - 87e (C2), 94 - 104 (D), 128 - 137 (E), 150 - 164 (F), 260 - 266 (G), 269 - 275 (H), 299 - 305 (I), 350 - 359 (R)
$\beta$ -sheet	142 - 145 (S1A), 110 - 121 (S2A), 182 - 196 (S3A), 331 - 340 (S5A), 291 - 298 (S6A), 228 - 232 (S1B), 237 - 243 (S2B), 248 - 255 (S3B), 370 - 376 (S4B), 382 - 388 (S5B), 50 - 52 (S6B), 363 - 365 (S1C), 281a - 289 (S2C), 215 - 227 (S3C), 204 - 209 (S4C)
$3_{10}$ -helix	200 - 202 (FI), 310 - 312 (FII)
$\beta$ -Turns - Type I	82 - 84, 165 - 168, 179 - 180, 233 - 236, 256a - 259, 313 - 316, 318 - 321, 324 - 327, 65 - 68, 87e - 87h
$\beta$ -Turns - Type II	45 - 48, 84 - 87, 91 - 94, 174 - 177, 244 - 246
$\beta$ -Turns - Type III	121 - 124, 277 - 280
$\alpha$ -Turn	148 - 150, 376 - 380

(the  $\alpha$ -turns are sharp turns that connect two  $\beta$ -strands)

Almost all of the polypeptide chain in ovalbumin is therefore involved in well defined structure. There are three main  $\beta$ -sheet regions (A - C), nine  $\alpha$ -helices (A - H, with helix R present in only one of the molecules) and three shorter helical segments (C2, FI and FII).

The backscattered Raman and ROA spectra of ovalbumin in H<sub>2</sub>O and D<sub>2</sub>O are shown in Figure 6.12.

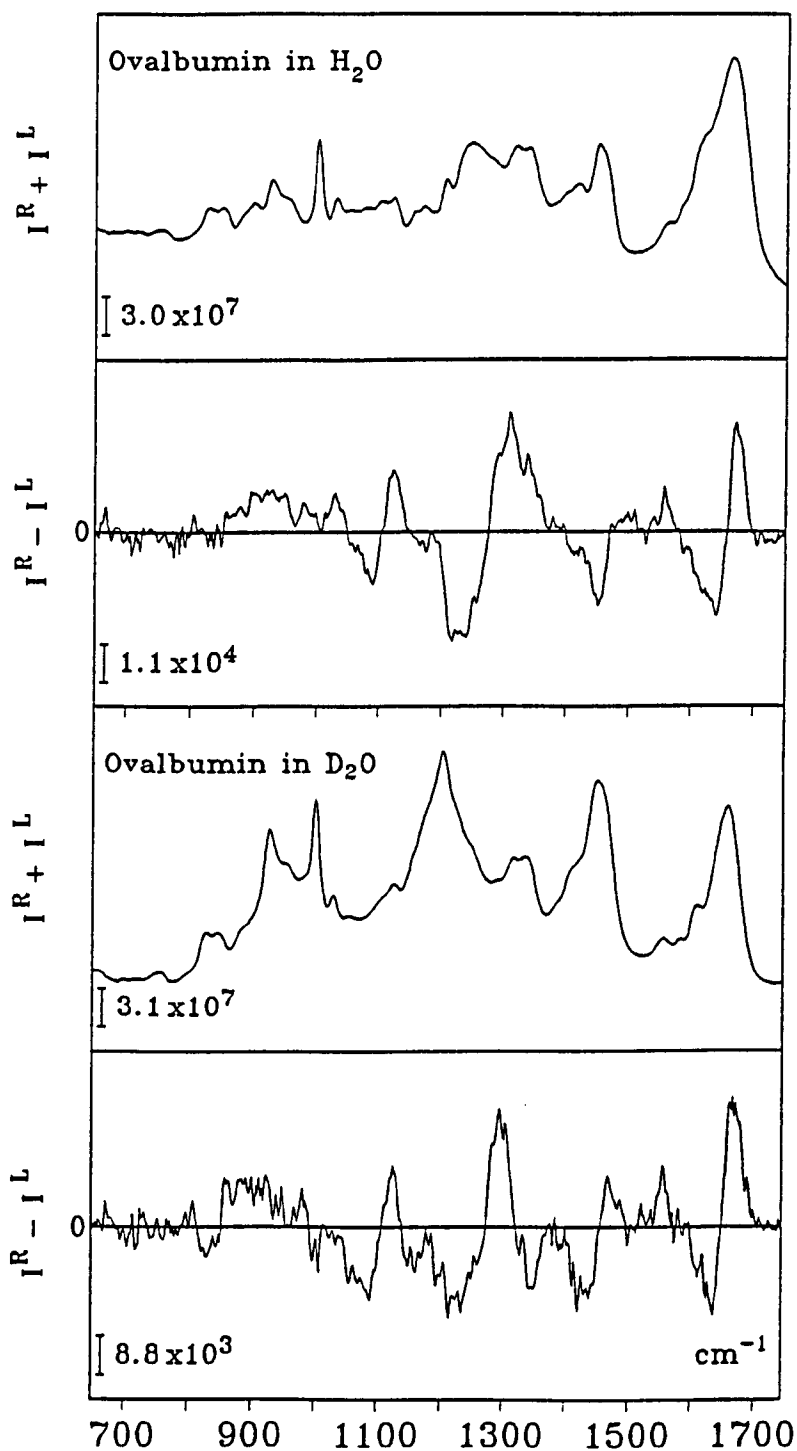


Figure 6.12 Backscattered Raman and ROA of ovalbumin



In the H<sub>2</sub>O ROA spectrum of ovalbumin the usual  $\alpha$ -helical signatures are present, namely the large positive ROA between  $\sim 851 - 963 \text{ cm}^{-1}$ ; the negative-positive couplet with a crossover at  $1103 \text{ cm}^{-1}$  and the underlying amide III couplet with a crossover at  $1273 \text{ cm}^{-1}$  are all indicators of a large amount of  $\alpha$ -helix being present. These three bands are also present in the D<sub>2</sub>O ROA spectrum as would be expected.

There is also a positive region between  $\sim 995 - 1046 \text{ cm}^{-1}$  and a sharp positive ROA band at  $1307 \text{ cm}^{-1}$ , which indicate the presence of  $\beta$ -sheet. A lot of the  $\beta$ -sheet is observed to exchange indicating that the sides of the exposed strands are subject to exchange. The usual negative-positive-negative pattern between  $\sim 950 - 1070 \text{ cm}^{-1}$  is not apparent here since the first negative part is 'masked' by a large positive region. It therefore cannot be said for certain from the ROA that the  $\beta$ -strands exist in a twisted environment.

In the H<sub>2</sub>O spectrum there is a sharp positive ROA band at  $1340 \text{ cm}^{-1}$ , again indicating the presence of  $3_{10}$ -helix, in agreement with the crystal structure (Stein *et al.*, 1991).

In D<sub>2</sub>O the  $\beta$ -turn signals become observable when they are not very clear in the H<sub>2</sub>O ROA spectrum. The sharp negative ROA band in the D<sub>2</sub>O spectrum at  $1348 \text{ cm}^{-1}$  indicates that there is a considerable amount of  $\beta$ -turns in the structure.

It is interesting that the two peaks at  $1212 \text{ cm}^{-1}$  and  $1237 \text{ cm}^{-1}$  in the bottom half of the amide III region which are some sort of loop structure, are not completely exchanged in D<sub>2</sub>O. These loop regions are therefore not completely exposed to solvent.

### $\beta$ -lactoglobulin

$\beta$ -lactoglobulin is abundant in the whey fraction of milk, but its actual function is as yet unknown. The protein has a molecular weight of  $\sim 18 \text{ kDa}$  and contains 162 amino acids. Although a crystal structure has been published for  $\beta$ -

lactoglobulin, it remains unclear as to why no coordinates have been deposited in protein databases (Papiz *et al.*, 1986). In Figure 6.13 there is a drawing of the  $\beta$ -lactoglobulin fold.

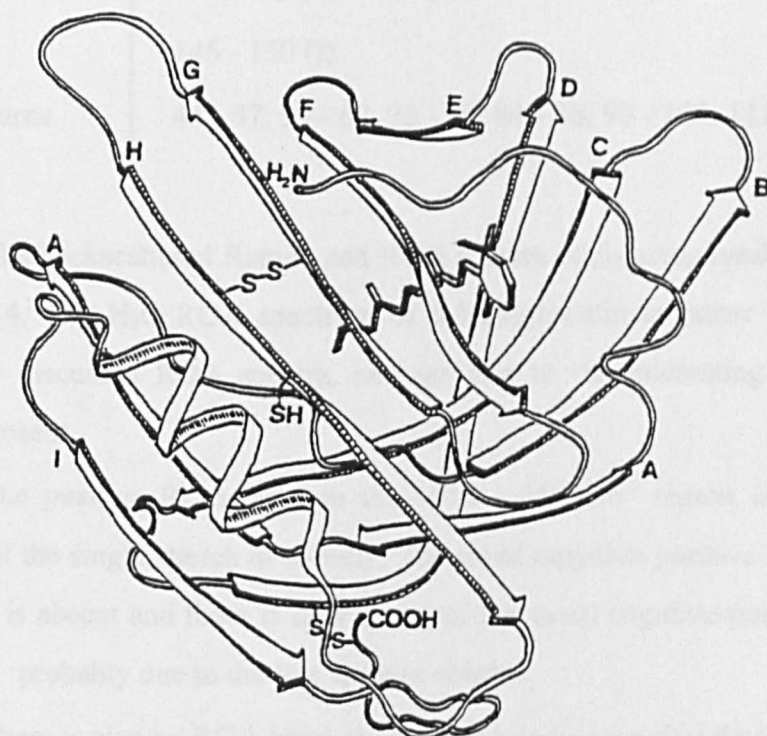


Figure 6.13 Representation of the  $\beta$ -lactoglobulin fold

The protein consists of nine strands in an antiparallel  $\beta$ -sheet arrangement that wrap round to form a flattened cone. In the core of the molecule there is an interesting structural motif, namely an eight-stranded antiparallel  $\beta$ -barrel (Branden & Tooze, 1991). This is also known as an up-down  $\beta$ -barrel which has been observed before in for example retinol-binding protein (Newcomer *et al.*, 1984) and catalase (Murthy *et al.*, 1981). Table 6.6 summarizes the structural elements present in  $\beta$ -lactoglobulin.

Table 6.6      Structure of  $\beta$ -lactoglobulin (Papiz *et al.*, 1986)

Structure	Residues
$\alpha$ -helix	133 - 140
$\beta$ -sheet	16 - 27 (A), 39 - 44 (B), 47 - 58 (C), 62 - 76 (D), 80 - 84 (E), 89 - 97 (F), 102 - 109 (G), 115 - 124 (H), 145 - 150 (I)
Turns	44 - 47, 59 - 62, 78 - 81, 84 - 88, 98 - 102, 111 - 115

The backscattered Raman and ROA spectra of  $\beta$ -lactoglobulin are shown in Figure 6.14. The  $\text{H}_2\text{O}$  ROA spectrum of  $\beta$ -lactoglobulin is rather different to the previously discussed ROA spectra, perhaps due to the interesting new structural features present.

The positive ROA band in the  $\sim 918 - 954 \text{ cm}^{-1}$  region is indicating the presence of the single stretch of  $\alpha$ -helix. The usual negative-positive ROA couplet at  $1103 \text{ cm}^{-1}$  is absent and there is only a hint of the usual negative-positive couplet at  $1275 \text{ cm}^{-1}$ , probably due to the low  $\alpha$ -helix content.

There is also no ROA band at  $1340 \text{ cm}^{-1}$ , indicating that there is no  $3_{10}$ -helix in the protein.

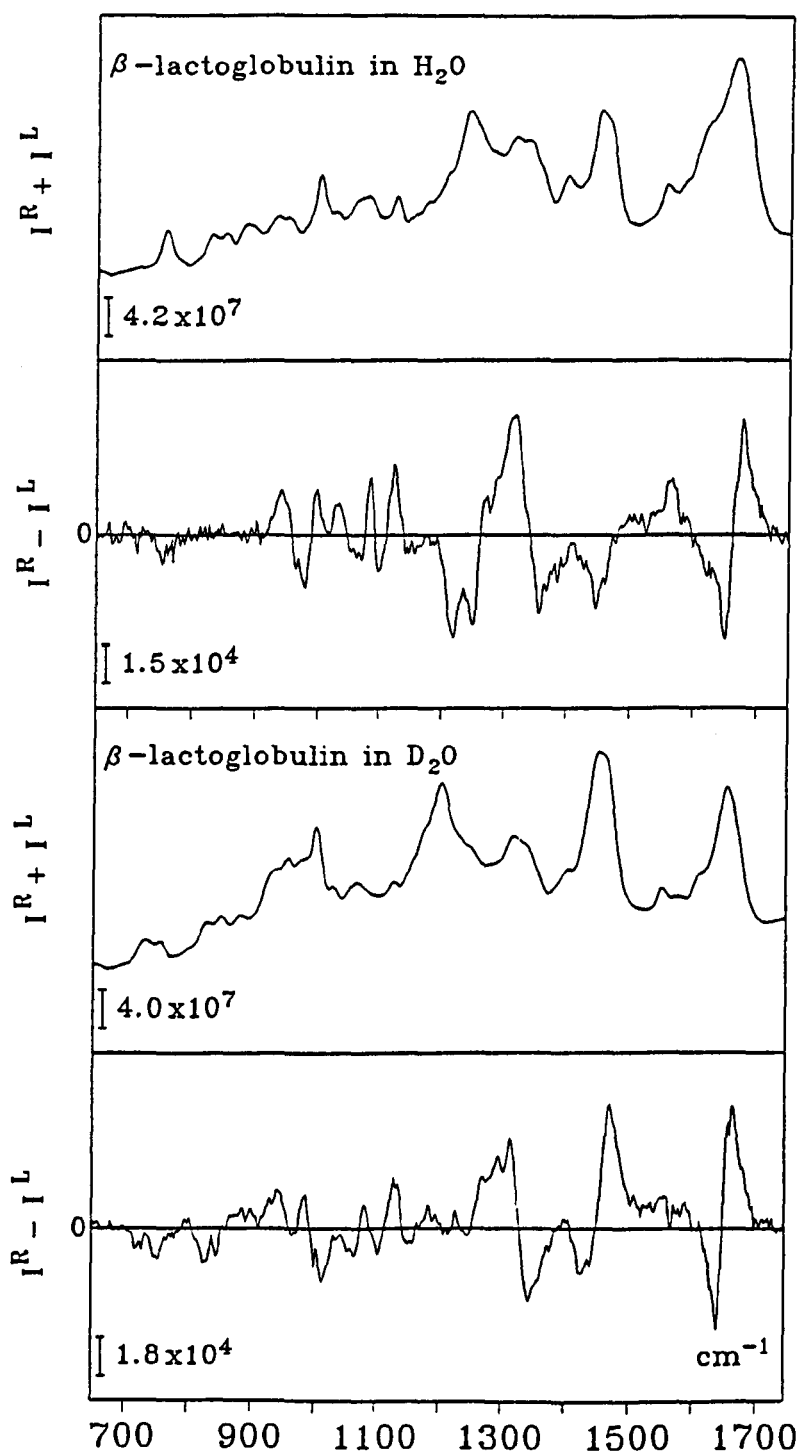


Figure 6.14 Backscattered Raman and ROA of  $\beta$ -lactoglobulin

The sharp ROA couplet with the negative part at  $1246\text{ cm}^{-1}$  and the positive part at  $1268\text{ cm}^{-1}$  has not been observed in any spectra so far (but it is observed in pepsin, see later). This ROA couplet might therefore be representative of the  $\beta$ - $\alpha$ - $\beta$  motif structure contained in  $\beta$ -lactoglobulin. If this is true this is a clear demonstration that ROA is recognising supersecondary structure. This theory could be tested by obtaining the ROA spectrum of retinol-binding protein (Newcomer *et al.*, 1984) which contains an antiparallel  $\beta$ -barrel structure similar to  $\beta$ -lactoglobulin but lacks the  $\beta$ - $\alpha$ - $\beta$  motif. In  $\text{D}_2\text{O}$  this band almost completely vanishes. It is again interesting that in the region of  $\sim 950 - 1070\text{ cm}^{-1}$  there is again a negative-positive-negative ROA pattern that is perhaps indicating that the  $\beta$ -strands are twisted.

Due to the high content of  $\beta$ -sheet there is a large ROA peak at  $1313\text{ cm}^{-1}$ . It is important to realize that due to the strands existing in a  $\beta$ -barrel, very few of the amide hydrogens will be exposed to exchange and thus the fact that the band persists in the  $\text{D}_2\text{O}$  ROA spectrum is giving information about the make-up of the  $\beta$ -sheet. Due to the high content of turn structure there is a sharp negative ROA band at  $1352\text{ cm}^{-1}$ , which is also found in the model  $\beta$ -turn structure. In  $\text{D}_2\text{O}$  there is again little effect on this band which is consistent with the assignment.

The amide I region has a sharp conservative couplet, with the negative part peaking at  $1647\text{ cm}^{-1}$  and the positive part at  $1674\text{ cm}^{-1}$ . A conservative couplet in the amide I region is therefore likely to be a signature for a high content of  $\beta$ -sheet.

### Concanavalin A

Concanavalin A is a member of the lectin family and is a carbohydrate binding protein. At physiological pH, concanavalin A usually exists as a tetramer, which is made up of four identical monomers of molecular weight  $\sim 25\text{ kDa}$  (Edmundson *et al.*, 1971; Wang *et al.*, 1971). The X-ray crystal structure has been solved (Becker *et al.*, 1975; Reeke *et al.*, 1975; Weisgerber & Helliwell, 1993) and it has been found that each monomer contains one saccharide binding site, and therefore four saccharides can be bound by the tetrameric concanavalin A (So &

Goldstein, 1968; Yariv *et al.*, 1968). An illustration of concanavalin A is shown in Figure 6.15.

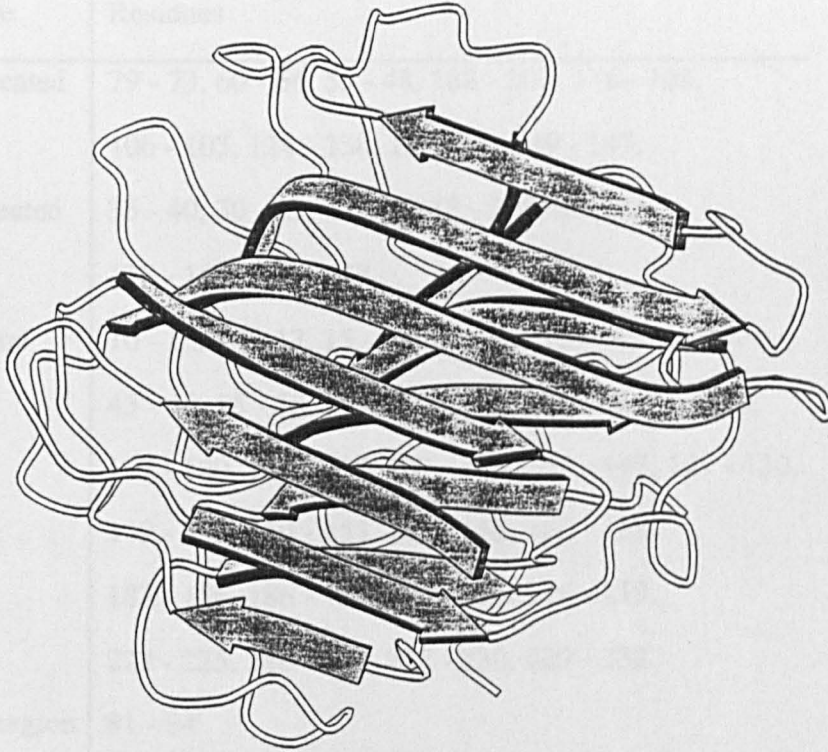


Figure 6.15 Molscript drawing of concanavalin A

The structure is constructed from a 'jelly roll' (Branden & Tooze, 1991) arrangement of the  $\beta$ -sheets. It consists of two large  $\beta$ -sheet structures: one forms the front of the molecule while the other runs through the molecular centre. Table 6.7 lists the structural features present in concanavalin A.

Table 6.7      Structure of concanavalin A (Weisgerber &amp; Helliwell, 1993).

Structure	Residues
Front pleated	79 - 73, 60 - 66, 55 - 48, 188 - 200, 116 - 108,
$\beta$ -sheet	106 - 103, 124 - 130, 153 - 16, 149 - 147.
Back pleated	35 - 40, 30 - 23, 4 - 11, 215 - 208, 88 - 97,
$\beta$ -sheet	176 - 169, 139 - 143.
Loops and turns	10 - 13, 14 - 17, 15 - 18, 28 - 31, 31 - 34, 34 - 37, 43 - 46, 55 - 58, 56 - 59, 67 - 70, 86 - 89, 97 - 100, 117 - 120, 134 - 137, 137 - 140, 143 - 147, 147 - 150, 149 - 152, 150 - 153, 160 - 163, 166 - 169, 183 - 186, 186 - 189, 203 - 206, 216 - 219, 222 - 225, 226 - 229, 227 - 230, 229 - 232.
Helical region	81 - 84

It can be observed from Table 6.7 that there is only a very small amount of helix present, namely residues 81 - 84 which form a single helical turn. There are a number of hairpin loops which allows the direction of the polypeptide backbone to change sharply.

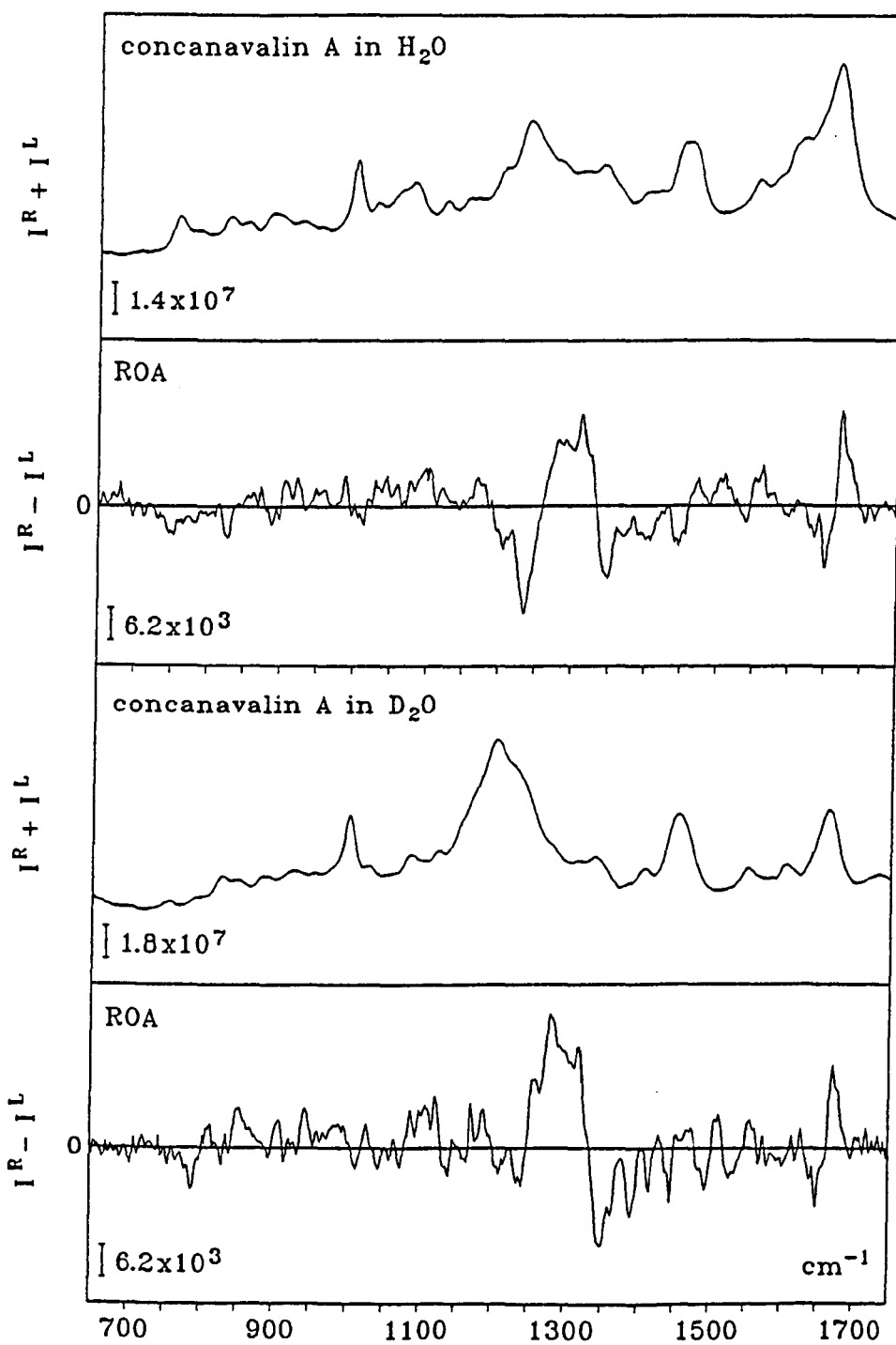


Figure 6.16 Backscattered Raman and ROA of concanavalin A



The backscattered Raman and ROA spectra of concanavalin A are shown in Figure 6.16 and are very informative about the structure of the protein. There is a sharp positive ROA band at  $1316\text{ cm}^{-1}$ , indicative of  $\beta$ -sheet, which only decreases slightly in  $\text{D}_2\text{O}$ , suggesting that the  $\beta$ -sheet structure is not in an open conformation. There is also the large negative ROA peak at  $1348\text{ cm}^{-1}$  which is characteristic of the many  $\beta$ -turns that are present in the structure.

The negative ROA band at  $1237\text{ cm}^{-1}$  which is characteristic of loop structure is also observed in the ROA of lysozyme. In  $\text{D}_2\text{O}$  this band almost completely vanishes.

In contrast to ribonuclease A which contains twisted  $\beta$ -sheets with the  $\beta$ -strand adopting a left-handed helical structure, there is no sign of the negative-positive-negative pattern in the  $\sim 950 - 1070\text{ cm}^{-1}$  region. This indicates that in solution the strands remain in the rigid  $\beta$ -barrel conformation and do not form twisted structures. It is also interesting that there is a lack of significant positive ROA intensity in the  $\sim 1020 - 1060\text{ cm}^{-1}$  region which was previously assumed to be due to antiparallel  $\beta$ -sheet (Wen *et al.*, 1994b). This region is perhaps therefore much more specific to the actual type of  $\beta$ -sheet structure than previously thought.

The amide I couplet is uncharacteristically sharp and may well have something to do with the rigid  $\beta$ -barrel conformation. The couplet has its negative part at  $1651\text{ cm}^{-1}$  and the positive part at  $1673\text{ cm}^{-1}$ . This couplet is virtually unchanged in  $\text{D}_2\text{O}$  apart from shifting down by  $10\text{ cm}^{-1}$ , indicating that some exchange has occurred.

## Pepsin

Pepsins are the major digestive enzymes designed to function at the low pH of the stomach. Pepsin itself consists of 326 residues with a molecular weight of ~33 kDa (Sepulveda *et al.*, 1975). The structure, which is shown in Figure 6.17, consists of two  $\beta$ -sheet lobes.

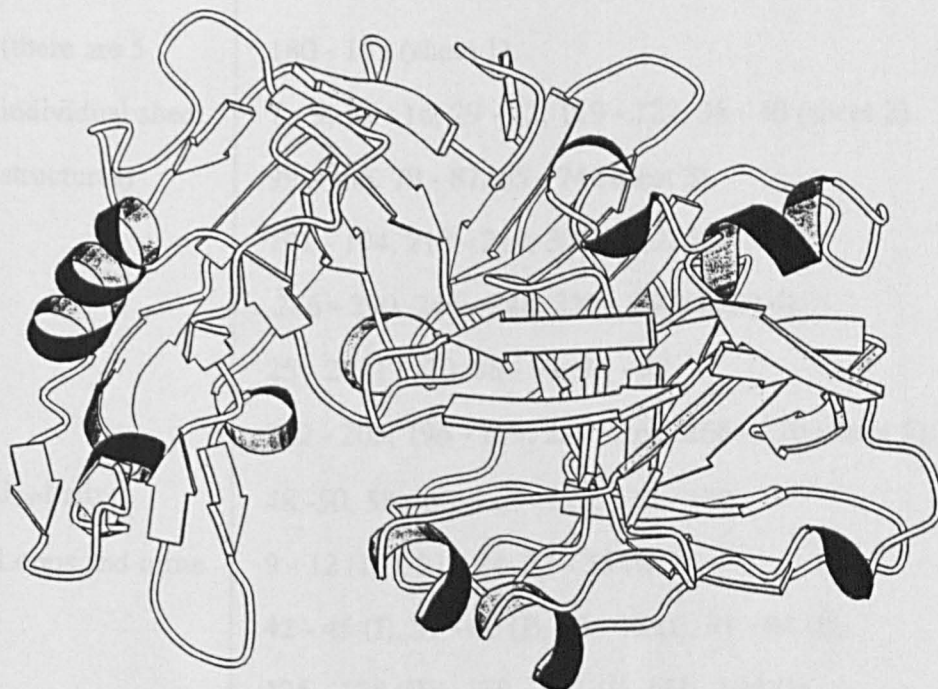


Figure 6.17 Molscript drawing of pepsin

It is found that the structure is dominated by  $\beta$ -sheet, with 44% of residues lying in  $\beta$ -strands (Kabsch & Sander, 1983; Cooper *et al.*, 1990; Sielecki *et al.*, 1990). Residues 1 - 175 form the N-terminal lobe and residues 176 - 327 comprise the C-terminal lobe.

Table 6.9      Summary of the structural elements in pepsin (Cooper *et al.*, 1990).

Structure	Residues
$\alpha$ -helix	110 - 114, 136 - 142, 225 - 236, 252 - 255, 272 - 275, 304 - 307.
$\beta$ -sheet (there are 5 individual sheet structures)	2 - 6, 163 - 167, 150 - 154, 310 - 315, 320 - 326, 180 - 183 (sheet 1). 7 - 9, 13 - 16, 29 - 32, 119 - 122, 38 - 40 (sheet 2). 99 - 106, 79 - 87, 65 - 74 (sheet 3). 191 - 194, 210 - 214, 300 - 302, 221 - 223, 286 - 290, 246 - 248, 239 - 240 (sheet 4). 25 - 28, 17 - 20, 88 - 91, 94 - 98, 202 - 205, 196 - 199, 259 - 263, 266 - 270 (sheet 5).
$3_{10}$ -helix	48 - 50, 58 - 60, 126 - 128, 130 - 132.
Loops and turns	9 - 12 (III), 21 - 24, 32 - 35 (I), 42 - 45 (I), 57 - 60 (I), 75 - 78 (I), 91 - 94 (I), 125 - 128 (III), 130 - 133 (I), 141 - 144 (I), 157 - 160 (III), 171 - 174 (I), 199 - 202 (I), 206 - 209 (II), 215 - 218 (I), 224 - 227 (III), 233 - 236 (I), 240 - 243 (I), 262 - 265 (I), 270 - 273 (III), 278 - 281 (I), 314 - 317.

The structure is therefore quite complicated with the central underlying sheet formed by six antiparallel  $\beta$ -strands.

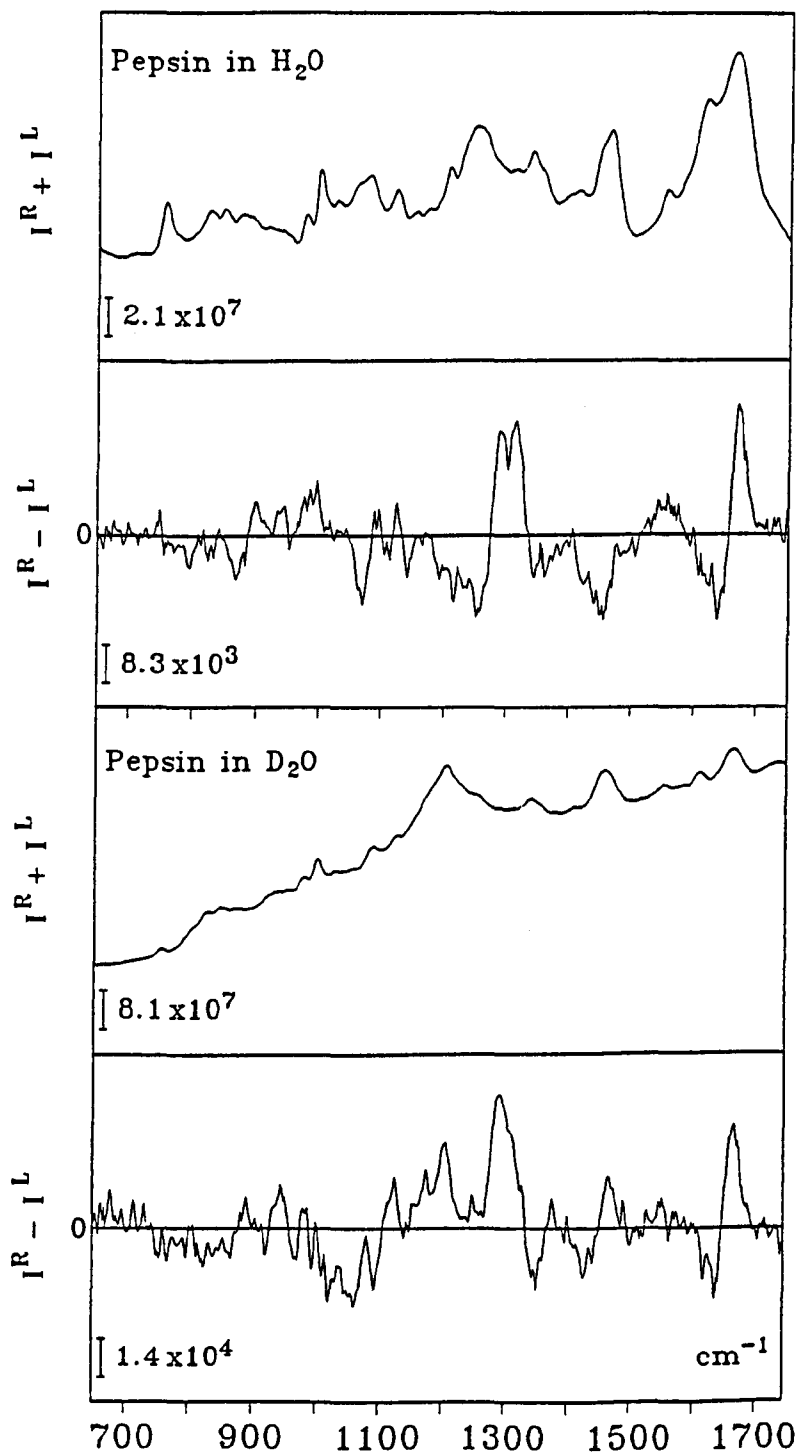


Figure 6.18 Backscattered Raman and ROA of pepsin

The backscattered Raman and ROA spectra of pepsin in H<sub>2</sub>O and D<sub>2</sub>O are shown in Figure 6.18. In the H<sub>2</sub>O ROA spectrum there is the usual positive ROA between  $\sim 890 - 940\text{ cm}^{-1}$  indicating the existence of  $\alpha$ -helix which surprisingly appears to be susceptible to deuterium exchange, although there are some positive peaks left in the D<sub>2</sub>O ROA spectrum (there is also the possibility that the ROA spectrum is slightly distorted). From the Molscript diagram the  $\alpha$ -helices are near the surface and might be part of the loop structure which would explain the high amount of deuterium exchange. Although the crystal structure indicates the presence of  $3_{10}$ -helix, there is no positive ROA band at  $1340\text{ cm}^{-1}$ . Perhaps in the solution the structure is much looser and not able to accommodate  $3_{10}$ -helix (or the negative  $\beta$ -turn ROA band cancels it out).

From the ROA the first impression is that the protein has a high  $\beta$ -sheet content. The sharp positive ROA band in the H<sub>2</sub>O spectrum at  $1311\text{ cm}^{-1}$  and the fairly conservative amide I negative-positive ROA couplet with its negative component peaking at  $1635\text{ cm}^{-1}$  and the positive part peaking at  $1668\text{ cm}^{-1}$ , suggest the presence of sheet structure. In D<sub>2</sub>O the amide I couplet is relatively unaffected but the  $1311\text{ cm}^{-1}$   $\beta$ -sheet band loses about half of its intensity. This indicates that the majority of the sides of the strands are susceptible to deuterium exchange. In the  $\sim 950 - 1070\text{ cm}^{-1}$  region there is again a negative-positive-negative ROA pattern indicating that the strands are perhaps twisted. There is also a sharp negative ROA band at  $1348\text{ cm}^{-1}$  which remains in D<sub>2</sub>O, generated by the large amount of  $\beta$ -turns present in the structure.

The sharp negative ROA band at  $1253\text{ cm}^{-1}$  which is completely lost in D<sub>2</sub>O, is generated by the loop structure in the pepsin. This is an unusual frequency for loop structure, not observed in lysozyme or lactalbumin. The entrance to the active cleft in pepsin is controlled by long mobile loops of  $\beta$ -strands protruding from both of the two lobes; these loops are capable of closing on a substrate and binding to it (Sugana *et al.*, 1987; Sielecki *et al.*, 1990). This seemingly unimportant piece of information is actually very interesting because a number of the proteins of the human immunodeficiency virus (HIV) are synthesized as single polypeptide chains

that an acid proteinase then splits into components. From the amino acid sequence it was assumed that the proteinase would have a similar structure to that of the acid proteinase of the pepsin family. From the crystal structure the proteinase was indeed found to be similar to that of pepsin (Navia *et al.*, 1989; Wlodawer *et al.*, 1989; Lapetto *et al.*, 1989). There could therefore be much promise for ROA in this area since it provides a new probe of the loop structure.

### $\alpha$ -chymotrypsin

$\alpha$ -chymotrypsin is a digestive enzyme and its biological role is to catalyze the hydrolysis of proteins in the small intestine. The protein contains 245 amino acids and has a molecular weight of ~25 kDa. The structure is shown in Figure 6.19, which reveals that the protein is made from two similar but not identical lobes.

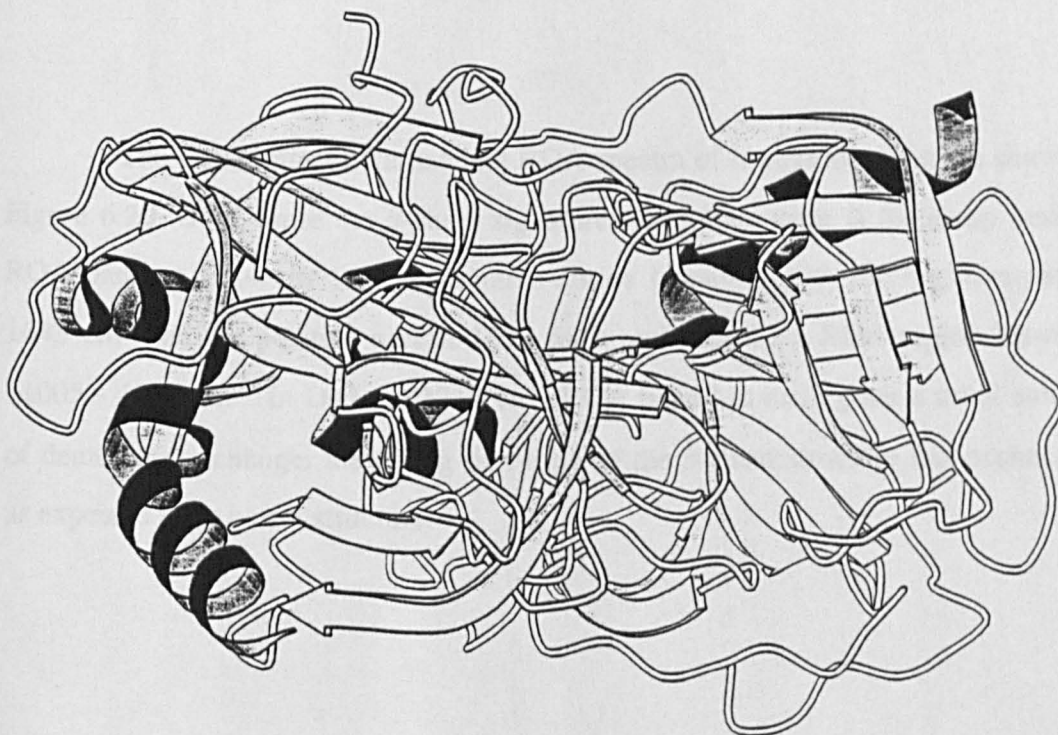


Figure 6.19 Molscript drawing of  $\alpha$ -chymotrypsin

$\alpha$ -chymotrypsin is again dominated by a  $\beta$ -barrel structure (Sigler & Blow, 1965; Tsukada & Blow, 1985) with a small amount of  $\alpha$ -helix. A summary of the secondary structure elements of one of the lobes of  $\alpha$ -chymotrypsin is listed in Table 6.9.

Table 6.9      Secondary structure of  $\alpha$ -chymotrypsin (Tsukada & Blow, 1985).

Structure	Residues
$\alpha$ -helix	164 - 176, 230 - 235, 235 - 245.
$\beta$ -sheet	29 - 35, 39 - 48, 51 - 54, 103 - 108, 81 - 91, 65 - 68, 29 - 35 (sheet 1). 134 - 140, 156 - 163, 179 - 184, 226 - 231, 206 - 215, 197 - 203, 134 - 140 (sheet 2).

The backscattered Raman and ROA spectra of  $\alpha$ -chymotrypsin are shown in Figure 6.20. Once more the  $\beta$ -sheet signatures are clear: there is the sharp positive ROA band at  $1314\text{ cm}^{-1}$ , a conservative amide I couplet with the negative part at  $1643\text{ cm}^{-1}$  and the positive part at  $1670\text{ cm}^{-1}$ ; and a positive ROA region between  $\sim 1005 - 1057\text{ cm}^{-1}$ . In  $\text{D}_2\text{O}$  the  $1314\text{ cm}^{-1}$  ROA band has undergone a small amount of deuterium exchange, indicating that little of the  $\beta$ -sheet structure has exchanged, as expected for a barrel structure.

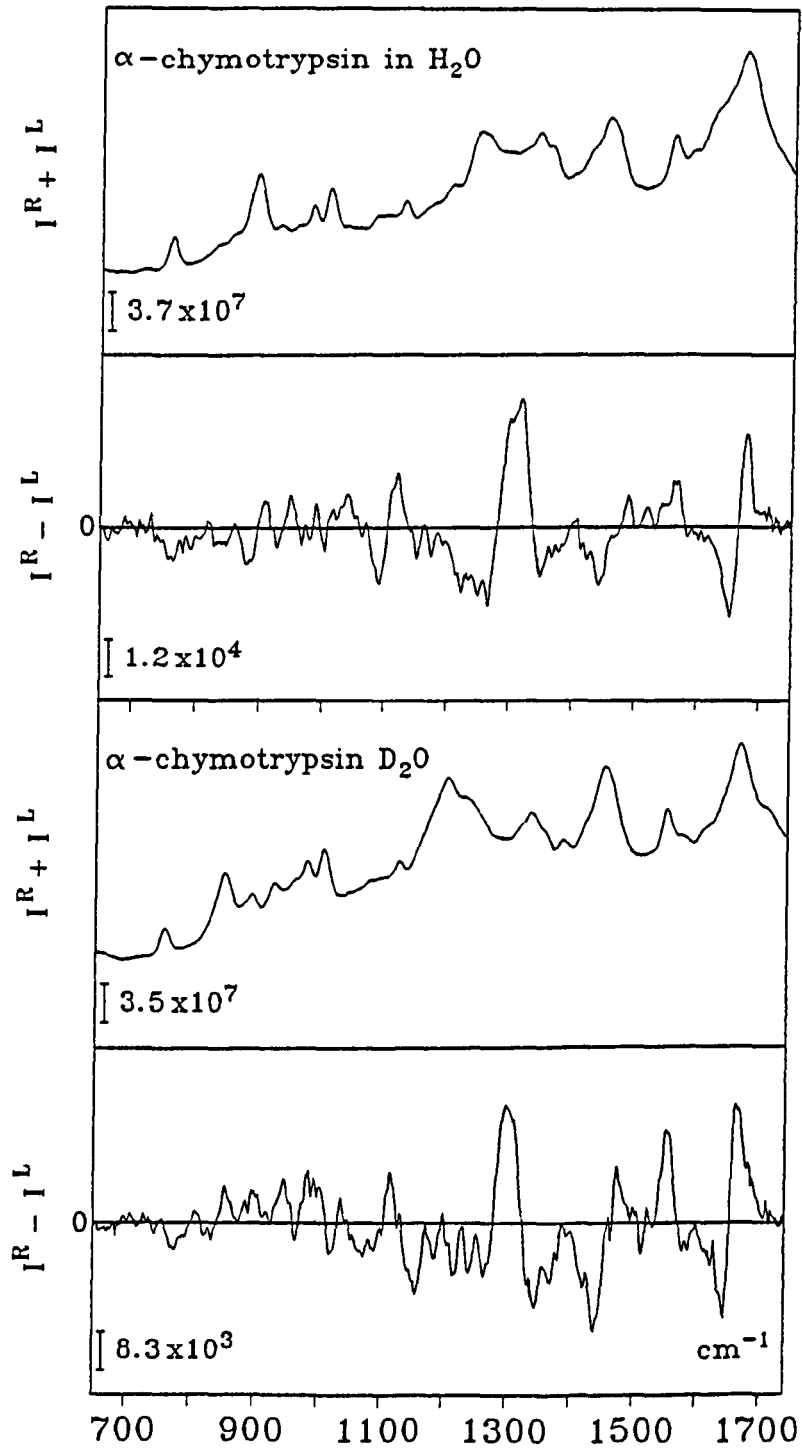


Figure 6.20 Backscattered Raman and ROA of  $\alpha$ -chymotrypsin



The negative feature observed at  $1348\text{ cm}^{-1}$  is a signature for the presence of  $\beta$ -turns. There also appears to be a lot of loop structure shown by the large amount of rather diffuse negative intensity in the lower wavenumber side of the amide III couplet. In  $\text{D}_2\text{O}$  the  $\beta$ -turn structure is more prominent and the loop region has undergone exchange but also appears to have become more structured.

There are also  $\alpha$ -helix signatures present, namely the positive ROA in the  $\sim 890 - 960\text{ cm}^{-1}$  region and the negative-positive ROA couplet with the crossover at  $1103\text{ cm}^{-1}$  which are relatively unaffected in  $\text{D}_2\text{O}$ .

## Trypsin

Trypsin is a digestive enzyme found in the stomach. Bovine trypsin contains 245 amino acids and has a molecular weight of  $\sim 23\text{ kDa}$ . The crystal structure has been solved by X-ray diffraction techniques (Bode & Schwager, 1975; Marquart *et al.*, 1983; Bartunik *et al.*, 1989). The structure of trypsin is shown in Figure 6.21.

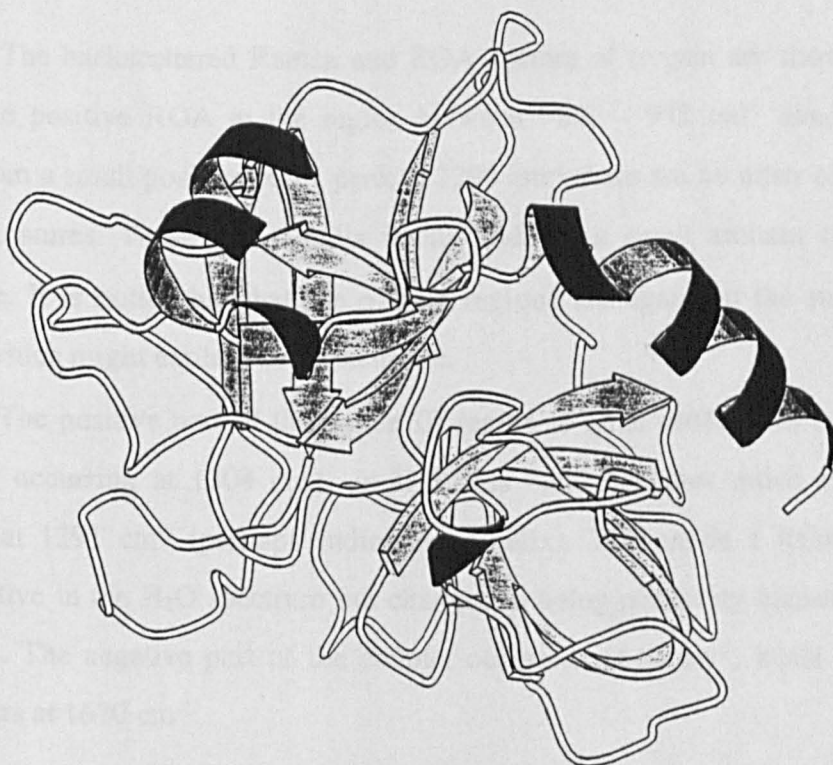


Figure 6.21 Molscript drawing of trypsin

The structure is a mixture of both  $\alpha$ -helix and  $\beta$ -sheet. The elements of secondary structure are listed in Table 6.10.

Table 6.10      Structure of trypsin (Bartunik *et al.*, 1989).

Structure	Residues
$\alpha$ -helix	56 - 58, 165 - 172, 173 - 175, 231 - 243, 242 - 244.
$\beta$ -sheet	30 - 33, 42 - 48, 51 - 54, 104 - 108, 81 - 90, 65 - 68, 30 - 33 (sheet 1). 135 - 140, 156 - 163, 180 - 183, 226 - 230, 207 - 215, 198 - 202, 135 - 140 (sheet 2).
Turns	23 - 26, 27 - 30, 33 - 41, 48 - 51, 72 - 76, 76 - 80, 91 - 94, 129 - 132, 145 - 149, 168 - 171, 177 - 179, 191 - 194, 194 - 197, 202 - 205, 222 - 224.

The backscattered Raman and ROA spectra of trypsin are shown in Figure 6.21. The positive ROA in the region between  $\sim 885 - 938 \text{ cm}^{-1}$  denotes  $\alpha$ -helix. Apart from a small positive ROA peak at  $1295 \text{ cm}^{-1}$  there are no other observable  $\alpha$ -helix signatures. These two  $\alpha$ -helix bands undergo a small amount of deuterium exchange. It is noticeable that the  $\alpha$ -helix regions are again on the surface of the protein which might explain their exchange.

The positive part of the amide III region is quite broad with the middle of the peak occurring at  $1304 \text{ cm}^{-1}$ . In  $\text{D}_2\text{O}$  this band becomes much sharper, now peaking at  $1291 \text{ cm}^{-1}$  (perhaps indicating  $\alpha$ -helix). The amide I ROA couplet is conservative in the  $\text{H}_2\text{O}$  spectrum but changes to being positively biased in the  $\text{D}_2\text{O}$  spectrum. The negative part of the couplet occurs at  $1640 \text{ cm}^{-1}$ , while the positive part occurs at  $1670 \text{ cm}^{-1}$ .

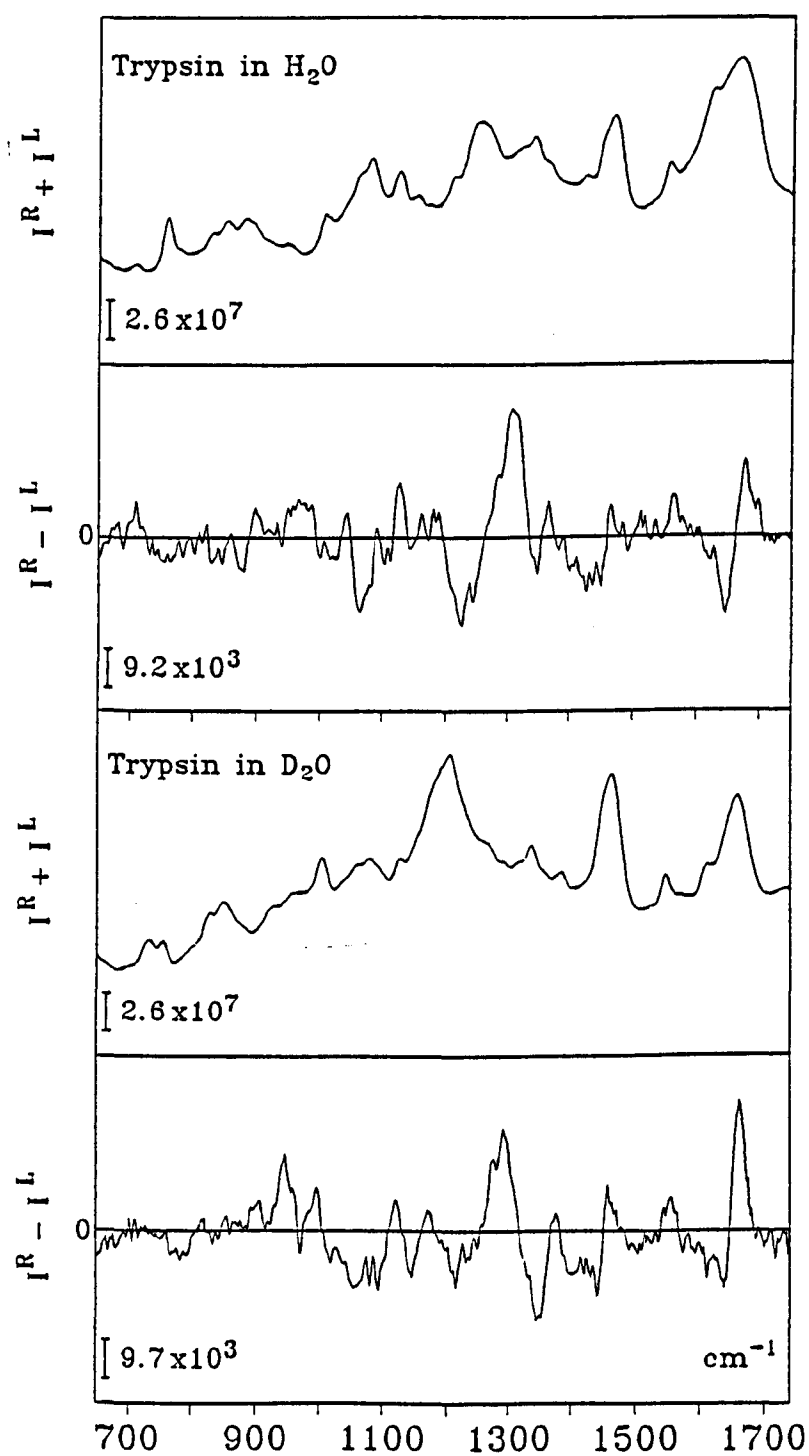


Figure 6.22 Backscattered Raman and ROA of trypsin

Due to the lack of any significant positive ROA intensity in the  $\sim 1000\text{ cm}^{-1}$  region and the lack of the usual negative-positive-negative ROA pattern between  $\sim 950 - 1070\text{ cm}^{-1}$ , it can be concluded that the  $\beta$ -strands are relatively flat in the solution structure.

There is also a sharp negative ROA band at  $1348\text{ cm}^{-1}$  both in  $\text{H}_2\text{O}$  and  $\text{D}_2\text{O}$  indicating that there is a considerable amount of turns present in the protein solution structure.

The two negative ROA loop bands at  $1223$  and  $1243\text{ cm}^{-1}$  are at exactly the same frequencies as the loop structure in  $\alpha$ -lactalbumin. In  $\text{D}_2\text{O}$  they are almost completely lost.

### Trypsinogen

Trypsinogen is a pancreatic zymogen which is used in the stomach to digest other proteins. The protein has a molecular weight of  $\sim 23\text{ kDa}$  containing 245 amino acids and contains a mixture of helix and sheet structure as shown in Figure 6.23 (Fehlhammer *et al.*, 1977; Kossiakoff *et al.*, 1977).

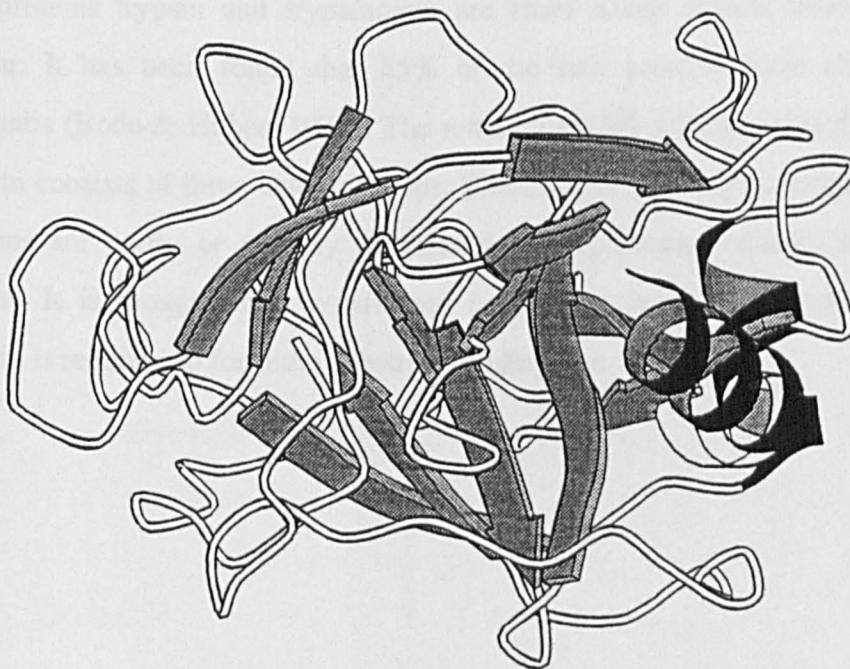


Figure 6.23 Molscript drawing of trypsinogen

Table 6.11 is a summary of the structural elements in trypsinogen.

Table 6.11            Structure of trypsinogen (Fehlhammer *et al.*, 1977).

Structure	Residues
$\alpha$ -helix	164 - 172, 234 - 245
$\beta$ -sheet	20 - 20, 156 - 161, 136 - 140, 197 - 201, 204 - 214, 226 - 231, 179 - 183 (sheet 1) 43 - 45, 52 - 55, 104 - 107, 87 - 91 (sheet 2) 63 - 66, 81 - 85 (sheet 3)
$3_{10}$ -helix	230 - 235
Turns	23 - 26, 48 - 51, 56 - 59, 91 - 94, 99 - 102, 115 - 118, 130 - 134, 172 - 175, 177 - 180, 184 - 187, 191 - 194, 194 - 197, 201 - 204, 221 - 224

The proteins trypsin and trypsinogen are from X-ray studies known to be very similar. It has been found that 85% of the two proteins have almost identical structures (Bode & Huber, 1978). The remaining 15% which is called the activation domain consists of three external loops. These loops, varying in length from 7 to 12 residues are partly or entirely disordered in trypsinogen crystals but ordered in trypsin. It is thought that trypsinogen is inactive because an ordered activation domain is required to form the substrate binding site.

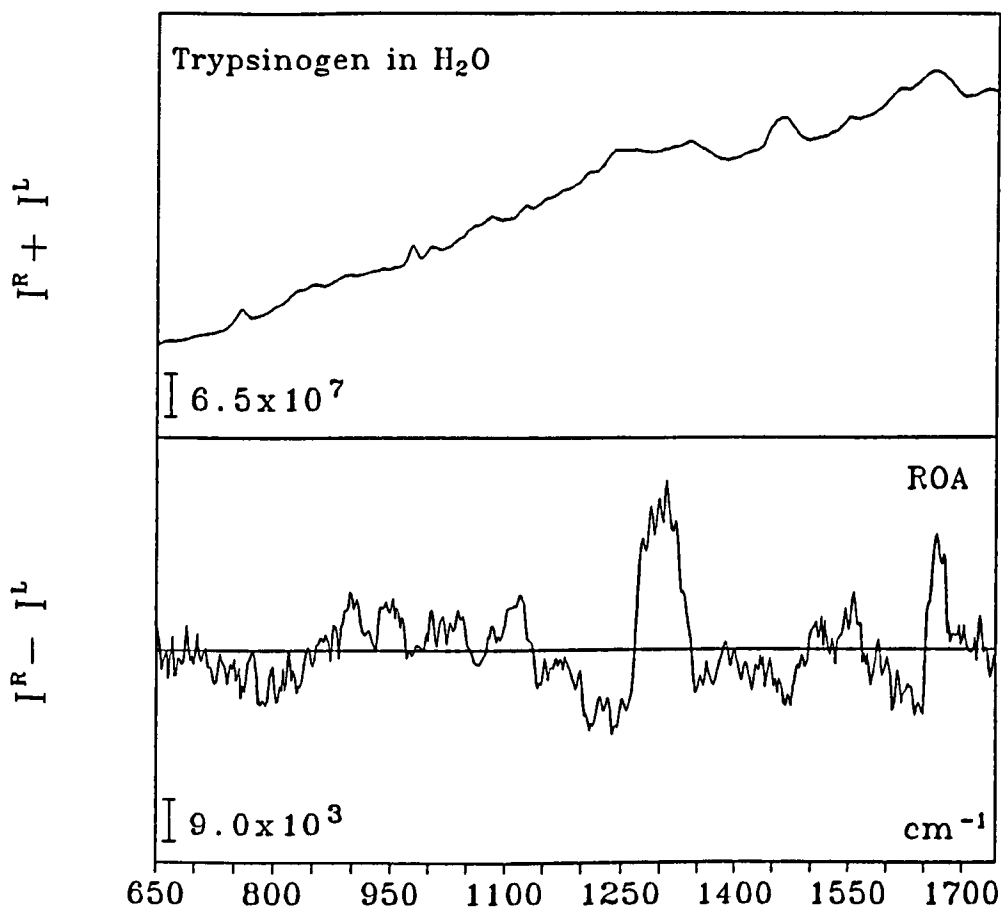


Figure 6.24 Backscattered Raman and ROA of trypsinogen

The backscattered Raman and ROA spectra of trypsinogen in  $\text{H}_2\text{O}$  are shown in Figure 6.24. It was unfortunate that the  $\text{D}_2\text{O}$  ROA spectrum of trypsinogen could not be obtained since the sample deteriorated too quickly in the beam to allow ROA to be obtained. Even the ROA spectrum in  $\text{H}_2\text{O}$  of trypsinogen is rather noisy.

The broad top of the amide III couplet peaks at about  $1304 \text{ cm}^{-1}$  and there is a broad positive region between  $\sim 995 - 1056 \text{ cm}^{-1}$  indicating the presence of  $\beta$ -sheet. There is a positive ROA region between  $\sim 890 - 940 \text{ cm}^{-1}$  which is representative of the  $\alpha$ -helix that is present in the structure. There is also a very small positive ROA peak at  $1340 \text{ cm}^{-1}$  indicating the presence of a very small amount of  $3_{10}$ -helix, which is in agreement with the crystal structure. In the  $\sim 950 - 1070 \text{ cm}^{-1}$  region there is the negative-positive-negative pattern which is an indication that there is not only  $\beta$ -sheet present but that the strands are twisted.

The sharp negative ROA peak at  $1348\text{ cm}^{-1}$  is due to the turn structure and the two peaks in the bottom half of the amide III at  $1212$  and  $1244\text{ cm}^{-1}$  are due to the loop structure. It is interesting that the amide III couplet is broader in trypsinogen than in trypsin, which is an indication that trypsin is the more rigid structure. Overall the ROA spectra of trypsin and trypsinogen are fairly similar as would be expected. It is also noticeable that the  $\sim 1223\text{ cm}^{-1}$  ROA band is much weaker than in trypsin and this could perhaps be providing information about the activation domain.

### Orosomuroid

Orosomuroid (or more commonly known as  $\alpha_1$ -acid glycoprotein) is a glycoprotein which is found in the blood and has a very high content of carbohydrate ( $\sim 40\%$ ). It is interesting to look at glycoproteins with ROA because they are very difficult to study by the usual physical methods such as X-ray crystallography, multi-dimensional NMR and ECD. Orosomuroid consists of 181 amino acids and has a molecular weight of  $\sim 41\text{ kDa}$ . The carbohydrate part involves five hetero-oligosaccharides which are attached by N-glycosidic bonds to asparagine residues (Schmidt *et al.*, 1973; Bell, 1994).

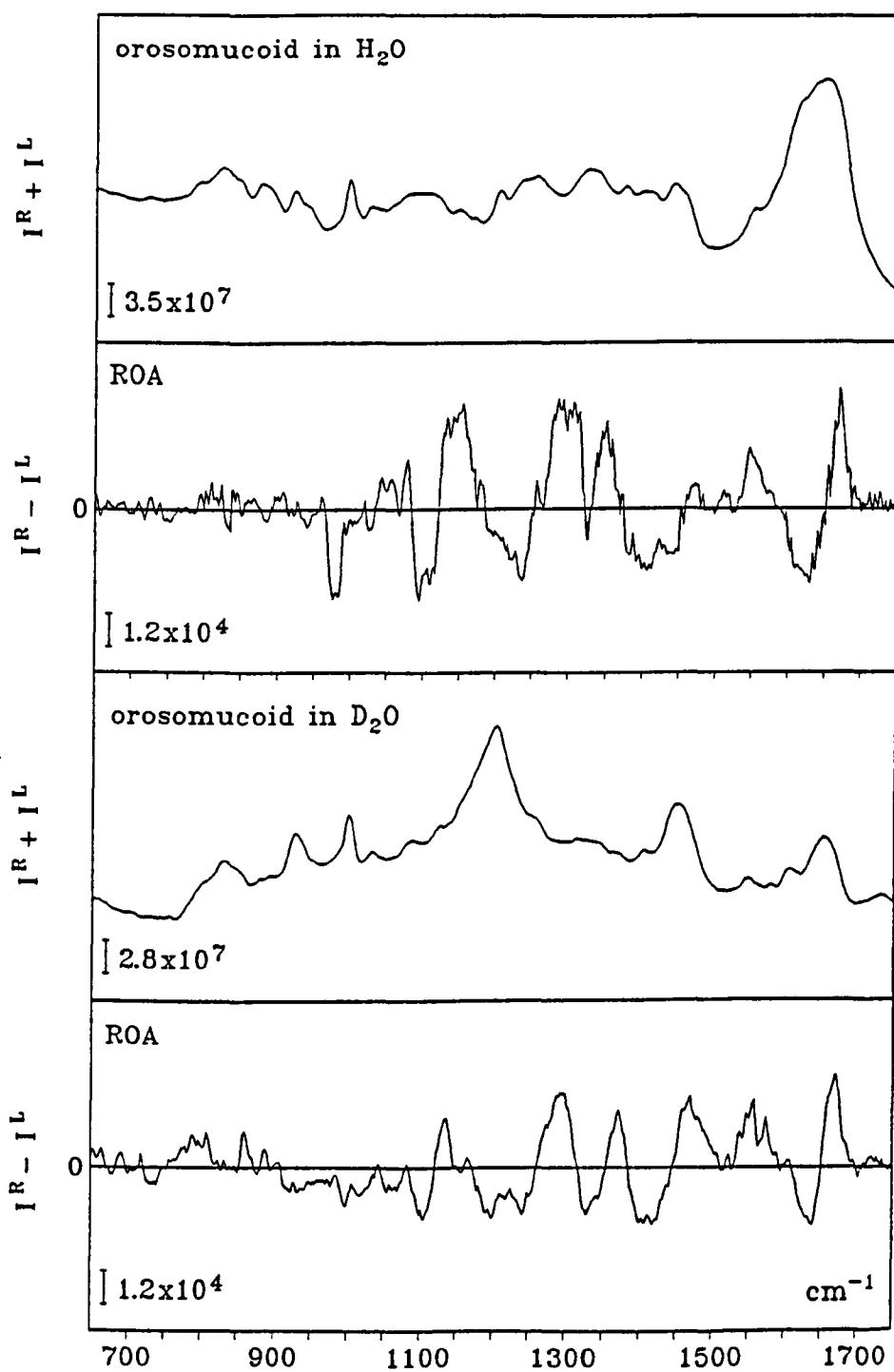


Figure 6.25 Backscattered Raman and ROA of orosomucoid



The backscattered Raman and ROA spectra of orosomuroid in both H<sub>2</sub>O and D<sub>2</sub>O are shown in Figure 6.25. The H<sub>2</sub>O ROA spectrum suggests that the glycoprotein is dominated by  $\beta$ -sheet structure. This observation conflicts with a recent study which concluded that there was only a small excess of  $\beta$ -sheet over  $\alpha$ -helix (Karpenko *et al.*, 1992).

The broad positive part of the amide III has one of its peaks at 1313 cm<sup>-1</sup> and the positive ROA between ~1030 - 1080 cm<sup>-1</sup>, both indicate that there is a significant amount of sheet present. The negative-positive-negative pattern in the ~950 - 1070 cm<sup>-1</sup> region also suggests that the sheets are twisted. It is a little surprising that there are no strong  $\alpha$ -helix bands present, apart from the amide III  $\alpha$ -helix couplet. The conservative amide I negative-positive couplet, with the negative part at 1629 cm<sup>-1</sup> and the positive part at 1671 cm<sup>-1</sup> also suggests that the structure is mainly  $\beta$ -sheet. It is also interesting that there is a sharp negative ROA band at 1237 cm<sup>-1</sup>, occurring in exactly the same position as the loop structure in lysozyme, which almost disappears in D<sub>2</sub>O.

There are however a few features observed in the ROA spectrum that have not been seen before. The negative-positive couplet centred at 1123 cm<sup>-1</sup> is in the region where previous IR studies (Marshall & Porath, 1965) of a glycopeptide of orosomuroid were found to have contributions from the carbohydrate. Previous disaccharide ROA studies (Bell *et al.*, 1993) suggested that this couplet could be due to the C-O-C stretching vibrations of the  $\beta$ (1-4) glycosidic link.

Future studies in this important area could involve the use of linkage-specific endoglycosidase enzymes to cleave the glycans from the polypeptide. The ROA could then be used to compare the carbohydrate part, the protein part and the complete glycoprotein.

## Conclusion

From the protein ROA spectra described in this chapter it is clear that ROA provides a unique insight into the structure and fold of proteins. It is important to realize that ROA not only provides information about the secondary but also tertiary structure. A summary of the ROA structural signatures will be made in the appendix.

## Chapter 7

### Protein Dynamics

In this chapter the reduced (i.e. unfolded) form of lysozyme and ribonuclease A, and molten globule states of  $\alpha$ -lactalbumin will be discussed. Molten globules are of great current interest because they are thought to be intermediates on the protein folding pathway and could provide a solution to the Levinthal paradox (Dobson, 1994). The Levinthal paradox suggests that it would take an almost infinite amount of time for a protein to fold properly if it had to explore all possible conformations. Interest in non-native states is rapidly increasing due to the importance of these states in protein folding, stability and function (Kim & Baldwin, 1990; Dill & Shortle, 1991; Creighton, 1992; Shortle, 1993; Pain, 1994; Dill *et al.*, 1995). Due to the heterogeneity associated with such states they have proved difficult to characterize with conventional methods such as NMR and far-UV CD.

The classic example of a molten globule state is supported by  $\alpha$ -lactalbumin at low pH or when the  $\text{Ca}^{2+}$  is extracted from the double loop region at neutral pH. A standard definition of a molten globule is that it is a compact structure that has native-like secondary structure but unfolded tertiary structure, the end result being that tertiary structure is much more mobile in the molten globule state than in the native state (Kuwajima, 1989). By varying the temperature of the molten globule state new insights into the possibility of native-like tertiary folds existing along with secondary structure (Ptitsyn, 1992) have been revealed by ROA.

The result of binding a saccharide inhibitor to lysozyme will also be studied. Finally, native lysozyme ROA spectra over a range of temperatures will be presented. The changes in these spectra will be discussed in relation to the structure and dynamics of Lysozyme.

The unfolded protein work and molten globule analysis is again an extension of the work of S.J. Ford (1995).

## 7.1 Experimental

Hen egg-white lysozyme (type I) and bovine  $\alpha$ -lactalbumin (type I- containing 1 to 2 mol  $\text{Ca}^{2+}$ /mol protein) were obtained from Sigma. Bovine  $\alpha$ -lactalbumin, which is partially calcium depleted (type III- containing about 0.4 mol  $\text{Ca}^{2+}$ /mol protein) was also purchased from Sigma. The bovine ribonuclease A was purchased from Fluka. All other materials including tri-N-acetyl glucosamine ( $\text{NAG}_3$ ),  $\text{GdHCl}$  and dithiothreitol (DTT) were again purchased from Sigma.

Unfolded lysozyme was prepared by reducing the four disulfide bridges using the method described by Creighton (Creighton, 1989). Briefly, this involved adding ~250 mg of protein to a 6M  $\text{GdHCl}$  solution at pH 8.0. When the protein had dissolved ~0.116 g of DTT was added. This solution was then stirred for 30 minutes. The reaction was then stopped by adding HCl until a pH < 1 was obtained. The purification stage was slightly altered from Creighton's method in that six 3 l volumes of 0.01 M HCl were used for dialysis of the sample, which was finally lyophilized. This proved to be an ideal method of denaturation for our purposes since using materials such as urea produced samples that fluoresced in the beam, making them unsuitable for ROA studies. Thermally denatured proteins are also unsuitable because these samples are highly scattering, due to aggregation. The broken disulfide bonds can reform and a way to block this is to carboxymethylate these residues to prevent reoxidation. However, the carboxymethylated lysozyme was insufficiently soluble for ROA studies. The procedure that was finally adopted was to prevent reoxidation by making the sample up at pH 2.0 in 10 mM citrate buffer, which was recently employed in an NMR study of denatured lysozyme (Evans *et al.*, 1991). An Ellman assay (Creighton, 1989) was used to check that the cysteine bonds were broken and this analysis showed that between seven and eight of the sulfhydryl groups were free. A similar procedure was employed to prepare reduced ribonuclease A except that a citrate buffer at pH 2.6 was used because this sample tends to form a gel at lower pH. Once the reduced samples for ROA study were prepared, the ROA spectra were obtained as quickly as possible to reduce any problems associated with

any refolding that might occur.

To fully remove the  $\text{Ca}^{2+}$  for studies of apo  $\alpha$ -lactalbumin ~100 mg of type III  $\alpha$ -lactalbumin was added slowly to 5 ml of a 20 mM EDTA/ $\text{NH}_3$  solution which was at pH 8.2. This solution was then stirred for 1 hour at room temperature. To ensure that as much of the  $\text{Ca}^{2+}$  as possible was removed, this was then dialysed against 5 x 200 ml volumes of EDTA/ammonia. The sample was then dialysed seven times against 3 l of filtered/deionised water to create as pure and clean a sample as possible before the sample was lyophilized.

Acid molten globule  $\alpha$ -lactalbumin was prepared in a 0.1 M glycine buffer at pH 2.0 buffer and the  $\text{Ca}^{2+}$  free  $\alpha$ -lactalbumin was prepared in a 0.1 M Tris buffer at pH 8.0 (it was found not to be sufficiently soluble at pH 5.4). To have a direct comparison native  $\alpha$ -lactalbumin was also prepared in a 0.1 M Tris buffer at pH 8.0.

The lysozyme bound to the trimer of N-acetylglucosamine ( $\text{NAG}_3$ ) was prepared by using a 2 molar excess of the saccharide inhibitor since it is known that at least a 1.3 molar excess is required to ensure that the lysozyme is fully bound (Lumb *et al.*, 1994). Due to  $\text{NAG}_3$  having the ability to hydrolyse the lysozyme, the ROA data was obtained as quickly as possible after preparing the sample in an acetate buffer at pH 5.4. The native lysozyme sample was prepared as explained previously in a 0.1 M acetate buffer.

Analar grade reagents and distilled, deionized water were used throughout. Using UV absorbance the protein concentrations of the  $\alpha$ -lactalbumin and lysozyme samples were determined. By measuring the absorbance of diluted aliquots and using an extinction coefficient for  $\alpha$ -lactalbumin of  $E_{280\text{nm}} = 2.01 \text{ ml}$  (Gill & von Hippel, 1989) and lysozyme of  $E_{280\text{nm}} = 2.65 \text{ ml}$ , concentrations were found to be about ~70 mg/ml. The protein solutions are prepared in small glass sample tubes and mixed with a small amount of pharmaceutical grade activated charcoal which helps to remove fluorescing impurities. The charcoal was removed from the sample by slow centrifugation. The protein solutions were then filtered through Millipore GV4 (0.22  $\mu\text{m}$ ) filters into quartz microfluorescence cells and centrifuged again to remove any dust particles prior to ROA data collection. All of these samples displayed residual

fluorescence but by leaving the sample to "burn down" in the laser beam for several hours the fluorescing impurities are selectively destroyed. This is a standard technique in the Raman spectroscopy of proteins (Carey, 1982) and does not appear to harm the sample as the Raman spectra remain unchanged throughout the procedure. On analysing the enzymatic activity of lysozyme (using the standard Sigma procedure based on the cell wall hydrolysis of *Micrococcus lysodeikticus*) and the fluorescence of the calcium-bound and calcium-free samples before and after ROA data collection confidence was gained that the protein molecules had been unaffected by the laser beam.

The tryptophan fluorescence measurements on diluted protein samples (0.05 to 0.1 mg/ml) over the temperature range 0°C to 35°C in a 1 cm path-length quartz cell were performed at a  $\lambda_{\text{exc}} = 295$  nm using a SPEX Fluor Max spectrofluorimeter. All of the fluorescence data were collected by Dr. A. Cooper. The UV CD measurements were performed on a Jasco J-600 spectropolarimeter using cells of path-length 0.02 cm (far-UV, concentrations ~0.6 mg/ml) and 0.05 cm or 0.5 cm (near-UV, concentrations ~6 mg/ml or ~0.6 mg/ml). The UV CD spectra were kindly measured by Prof. N.C. Price and Dr. S.N. Kelly at the BBSRC funded facility at Stirling University.

## 7.2 Results and Discussion

The ROA spectra will now be discussed in the context of protein structure and dynamics.

### Unfolded Lysozyme and Ribonuclease A

The backscattered Raman and ROA spectra of reduced lysozyme at 45°C, 20°C and 2°C are shown in Figure 7.1.

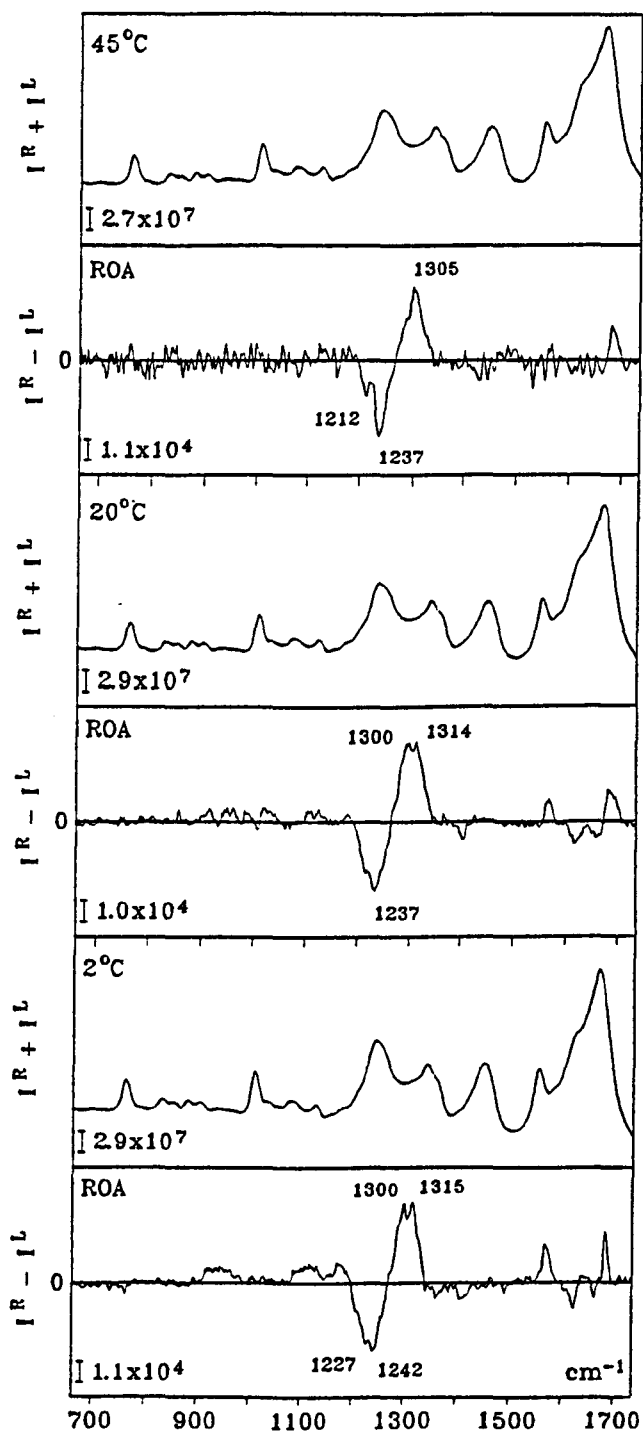


Figure 7.1 Backscattered Raman and ROA of unfolded lysozyme at 45, 20 and 2°C

Comparing these spectra with that of native lysozyme (Figure 6.4) there are obvious differences in the spectra, especially since the ROA spectra in the unfolded protein is now dominated by a relatively unstructured extended amide III couplet.

The residual structure that does exist in the unfolded proteins is nonetheless of great significance since it appears to monitor the  $\phi$ ,  $\psi$  propensities of the individual residues averaged over the different amino acids in the mobile heteropolypeptide structure. In the 20°C ROA spectrum there are two positive ROA bands at 1300 and 1313  $\text{cm}^{-1}$ , which are close to previously mentioned bands that have been assigned to  $\alpha$ -helix and  $\beta$ -sheet, respectively, in native proteins (see Chapter 6). The negative ROA band at 1237  $\text{cm}^{-1}$  in the 20°C spectrum coincides with an ROA band in lysozyme which was previously assigned as 'loop' structure. It is intriguing that in recent work on protein architecture (Adzhubei & Sternberg, 1993), analysis of all of the  $\phi$ ,  $\psi$  angles in a set of 68 proteins from the crystallographic structural database (Adzhubei *et al.*, 1987) revealed three main  $\phi$ ,  $\psi$  distribution peaks. Adzhubei *et al.*, (1987) have associated these peaks with  $\alpha$ -helix,  $\beta$ -structure and  $M$ -conformation, the last corresponding to PPII type structure. Since the  $\alpha$ -helix and  $\beta$ -sheet regions have already been fairly well established in this loop region, perhaps the negative 1237 $\text{cm}^{-1}$  ROA band is derived from PPII type structure. However, this is a very tentative assignment which requires much more work to fully validate this idea.

The extra small bumps in the extended amide III region are presumably due to residues populating less favourable regions of  $\phi$ ,  $\psi$  space.

In the 2°C ROA spectrum the previously discussed negative ROA band at 1237  $\text{cm}^{-1}$  has split into two components, one at 1227  $\text{cm}^{-1}$  and the other one at 1242  $\text{cm}^{-1}$  which suggests that at the lower temperature structure with distinct but related types of local order are becoming resolved in the putative PPII-type structure. On the other hand in the 45°C ROA spectrum the  $\alpha$ -helix and  $\beta$ -structure ROA bands at 1300 and 1314  $\text{cm}^{-1}$  have coalesced into a single ROA band at 1305  $\text{cm}^{-1}$ , indicating the loss of the two local residue conformations. The 1237  $\text{cm}^{-1}$  negative ROA band has also become sharper and is now accompanied by another negative ROA band at



1212  $\text{cm}^{-1}$ . The significance of these changes is as yet unknown.

An analogy of what has happened in the temperature-dependent ROA spectra in the amide III region of lysozyme can be found in chemical exchange effects in NMR where closely spaced lines collapse into a single line at the temperature at which the lifetime broadening is greater than the line separation. The timescale for ROA can be obtained from the timescale for conventional Raman scattering ( $\sim 3.3 \times 10^{-14}$  s for a vibration of  $\sim 1000$   $\text{cm}^{-1}$ ). This therefore represents a very fast timescale with ROA in effect providing a superposition of 'snapshot' spectra from all the distinct chiral conformers. At 45°C the ROA could therefore be undergoing broadening due to the finite lifetimes of the interconverting conformers. This effect has been observed for conventional vibrational Raman lines (Kreevoy & Mead, 1965). There has also been recent IR work along these lines (Turner *et al.*, 1995). From the spectra and using the formula  $k \approx 2.2\delta\nu$  an approximate rate constant for the temperature at which two bands with frequency separation  $\delta\nu$  coalesce can be found (Günther, 1995). From the ROA spectra the separation of the  $\alpha$ -helix and  $\beta$ -structure bands is  $\sim 14$   $\text{cm}^{-1}$  and therefore the rate constant at 45°C for interconversion between the corresponding conformers is  $\sim 9.2 \times 10^{11}$   $\text{s}^{-1}$ . This value is similar to that of the rate of rotations involving changes in  $\phi$  and  $\psi$  about  $\text{C}_\alpha\text{-N}$  and  $\text{C}_\alpha\text{-C}$  bonds, respectively, in single residues in the polypeptide backbone (Creighton, 1994). The simplest version of absolute reaction rate theory predicts a maximum rate for a chemical process of  $6.2 \times 10^{12}$   $\text{s}^{-1}$  at 25°C corresponding to a zero free energy barrier (Creighton, 1994). This value will be smaller when collision damping is included (Brooks *et al.*, 1988). The ROA spectra are therefore suggesting that, at physiological temperatures, the individual residues in mobile heteropolypeptide structure in unfolded lysozyme 'flicker' between distinct conformational states at rates close to the maximum theoretically possible. It is also worth remembering at this stage that the unfolded lysozyme is completely different from that of the low pH unordered poly-L-lysine (Figure 5.5).

It is interesting that in a recent study which excluded all residues in  $\alpha$ -helix or  $\beta$ -strand, three main regions of  $\phi$ ,  $\psi$  space were found to be present for individual

residues in 'coil' regions of native proteins (Swindels *et al.*, 1995). These regions were termed a, b and p but they appear to have a too diffuse range of  $\phi$ ,  $\psi$  angles to give rise to the sharp ROA peaks. This therefore suggests that in the crystal the coil regions of native proteins the residue conformations are distorted by the surrounding crystal packing forces, whereas in aqueous solution the  $\phi$ ,  $\psi$  propensities are much more tightly focused.

An important feature that is absent in the ROA spectra of unfolded lysozyme is the lack of a sharp positive band at  $1340\text{ cm}^{-1}$  which has previously been assigned to  $3_{10}$ -helix loop structure. It can be concluded therefore that there is little propensity for individual residues to take up this conformation, unlike PPII-helix loop structure.

The backbone skeletal stretch region of the unfolded lysozyme at  $20^\circ\text{C}$  indicates that there is only a very small amount of extended secondary structure present. On lowering the temperature to  $2^\circ\text{C}$  clearer  $\alpha$ -helix signals appear, namely the broad positive ROA band peaking at  $950\text{ cm}^{-1}$  and the positive component on the high wavenumber side of the couplet at  $1100\text{ cm}^{-1}$ . Comparing the unfolded  $2^\circ\text{C}$  spectrum with that of the native lysozyme ROA spectra (Figure 6.4) there is  $\sim 25\%$  of the native amount of  $\alpha$ -helix present. It is also noticeable that there is no twisted  $\beta$ -strand structure. The ROA couplet positive at  $1555\text{ cm}^{-1}$  and negative at  $1580\text{ cm}^{-1}$  which is due to the six tryptophans in the protein increases along with the  $\alpha$ -helix bands at  $2^\circ\text{C}$  but has almost disappeared at  $45^\circ\text{C}$ . This suggests that the conformational freedom of the tryptophans associated with sections of  $\alpha$ -helix has been much reduced. These regions could be due to helix C and D which encompass residues 88 - 99 and 108 - 115, respectively (Figure 6.1). A peptide containing residues 84 - 129 of lysozyme at pH 2.0 and  $25^\circ\text{C}$  was shown from a UV CD study to have a high intrinsic propensity for  $\alpha$ -helix formation (Yang *et al.*, 1995). Another VCD study for a peptide comprising residues 36 - 108 again found that residues corresponding to helix C had a high propensity for  $\alpha$ -helix formation (Ueda *et al.*, 1994). It should also be observed that at  $45^\circ\text{C}$  there is virtually no secondary structure present.

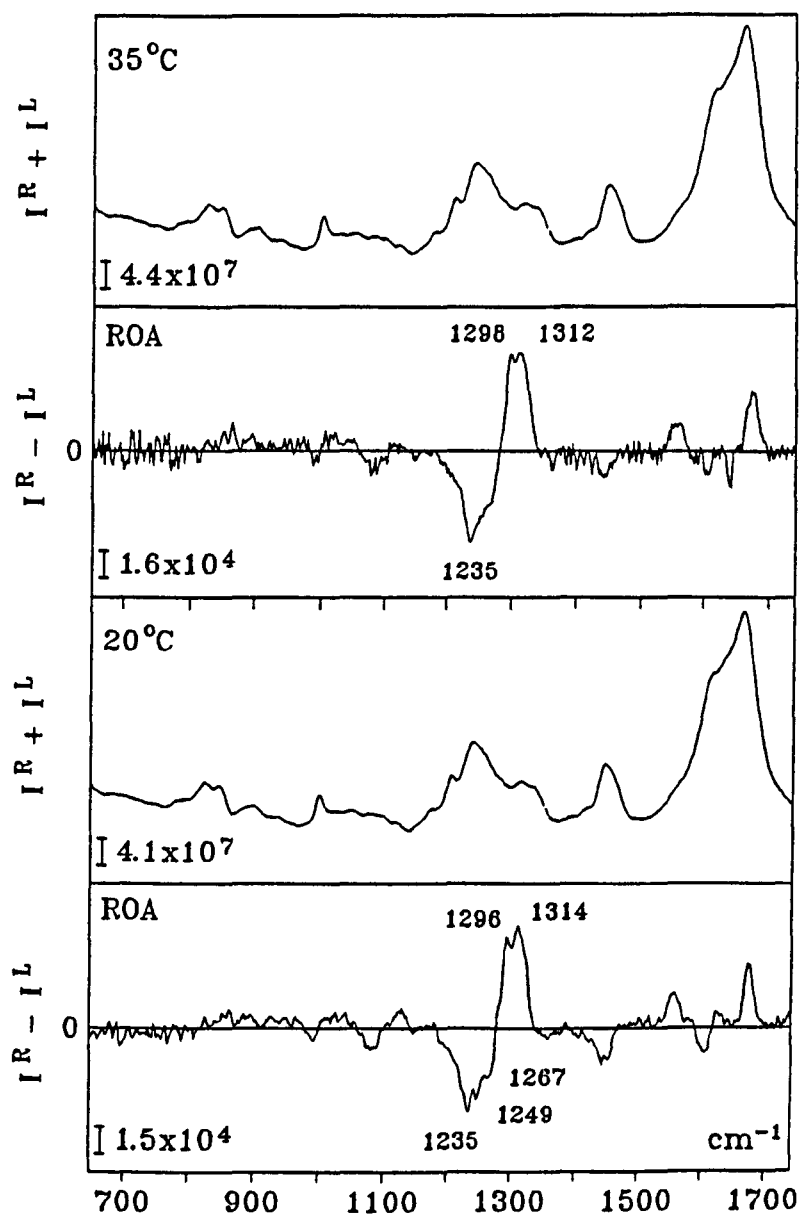


Figure 7.2 Backscattered Raman and ROA of unfolded ribonuclease A at 35 and 20°C

At 20°C and 45°C there is an amide I negative-positive ROA couplet that is shifted by  $\sim 10\text{ cm}^{-1}$  to a higher wavenumber than that found in the native protein. It is however much smaller than the native amide I couplet and presumably arises from the local chiral environments of the carbonyl oscillators.

The backscattered Raman and ROA spectra of unfolded ribonuclease A at 35°C and 20°C are shown in Figure 7.2. It was unfortunate not to be able to obtain ROA spectra over a wider temperature range but the unfolded ribonuclease A was found to form a gel and aggregate at higher or lower temperatures. Comparing the 20°C ROA spectra of unfolded lysozyme and ribonuclease A, it is clear that the unfolded ribonuclease A contains much more structure.

In the extended amide III region of the 20°C ROA spectrum the three bands assigned to individual structures containing  $\alpha$ -helix,  $\beta$ -structure and PPII type structure can be clearly observed. There are slight differences in the position of these bands though which is highly relevant. The  $\alpha$ -helix band is  $\sim 4\text{ cm}^{-1}$  lower in the ribonuclease A, which could be due to the greater amount of extended  $\alpha$ -helix secondary structure present ( $1296\text{ cm}^{-1}$  is in fact the usual value for the  $\alpha$ -helix band in native proteins). Raising the temperature to 35°C it was found that this band shifts by  $\sim 2\text{ cm}^{-1}$  to a higher wavenumber, with the ROA band now being of equivalent height to the adjacent  $\beta$ -structure band. In the ROA of unfolded ribonuclease A at 20°C (and to a lesser extent at 35°C) there is a negative ROA band at  $1360\text{ cm}^{-1}$ , indicating the presence of a small amount of turn structure.

It is also apparent in both the unfolded ribonuclease A spectra that there is no positive ROA band at  $1340\text{ cm}^{-1}$ , indicating again that there is no  $3_{10}$ -helix loop structure. At 35°C the extended amide III couplet is very smooth (unlike unfolded lysozyme) suggesting that there are few residues taking up  $\phi$ ,  $\psi$  angles other than those corresponding to the three main bands.

In the backbone skeletal stretch region of the unfolded ribonuclease A at 20°C in the range  $\sim 950 - 1150\text{ cm}^{-1}$  there is  $\sim 50\%$  of the native secondary structure. The remaining secondary structure also appears to have right-twisted  $\beta$ -strand, as does the native form. The  $\alpha$ -helix structure might occur at the amino-terminal end

since residual structure has been detected there by nonradiative transfer (Haas *et al.*, 1988). A UV CD analysis of a peptide of ribonuclease A comprising residues 1 - 13 also found the structure to be ~30% helical at 1.7°C (Brown & Klee, 1971). Helical conformations have also been detected in this peptide by 2-dimensional NMR (Osterhout *et al.*, 1989). At 35°C in the unfolded ribonuclease A although there is still secondary structure present, the amount has been greatly reduced.

Considering both the unfolded lysozyme and ribonuclease A it is apparent that there is no positive ROA band at 1340 cm<sup>-1</sup> and therefore it can be concluded that there is no  $3_{10}$ -helix loop structure, nor any propensity for individual residues to take up the associated  $\phi$ ,  $\psi$  angles in these two unfolded proteins and that a structured environment is required for this to form. This conclusion actually goes against that of Millhauser (1995) who suggested  $3_{10}$ -helix is a precursor to  $\alpha$ -helix in the protein folding pathway. It can also be suggested from this result that there is no interconversion between  $\alpha$ - and  $3_{10}$ -helix although other models have been built around this idea (Smythe *et al.*, 1995).

It is interesting to note that in the protein GBI an identical 11-amino sequence folds as  $\alpha$ -helix in one part of the protein but as  $\beta$ -sheet in another part (Minor & Kim, 1996). This suggests that the environment is of fundamental importance in determining the structure. In both the unfolded proteins presented here it can be observed that  $\alpha$ -helix and  $\beta$ -sheet can exist as extended structure which is probably due to the high propensities for individual residues to take up corresponding regions of  $\phi$ ,  $\psi$  space. It is perhaps not a coincidence that the unfolded lysozyme at 20°C is very similar to that of the unfolded ribonuclease A at 35°C despite both proteins having quite different amino acid sequences. This ties in with the result of Minor and Kim (1996) since it suggests that factors other than intrinsic residue propensities ultimately determine the actual conformation adopted by a particular stretch of the peptide backbone in a native protein.

### Acid Molten Globule $\alpha$ -Lactalbumin

The backscattered Raman and ROA spectra of the acid molten globule of  $\alpha$ -lactalbumin at 2°C, 20°C and 35°C are shown in Figure 7.3. As previously stated the  $\alpha$ -lactalbumin is at pH 2.0 in a glycine buffer and is in a so-called 'molten globule' state (Kuwajima, 1989; Ptitsyn, 1992), where the tertiary structure is presumed to be mobile yet retaining rigidity in some of the secondary structure. In Figure 7.4 a complete range of ROA spectra at 45°C, 35°C, 25°C, 20°C, 15°C, 10°C, 8°C and 2°C are stacked on top of each other. By collecting the ROA over a range of temperatures it might be possible to separately characterise transitions involving secondary and tertiary loop structure which could be of importance for protein folding since the molten globule is thought to be a universal intermediate on the folding pathway.

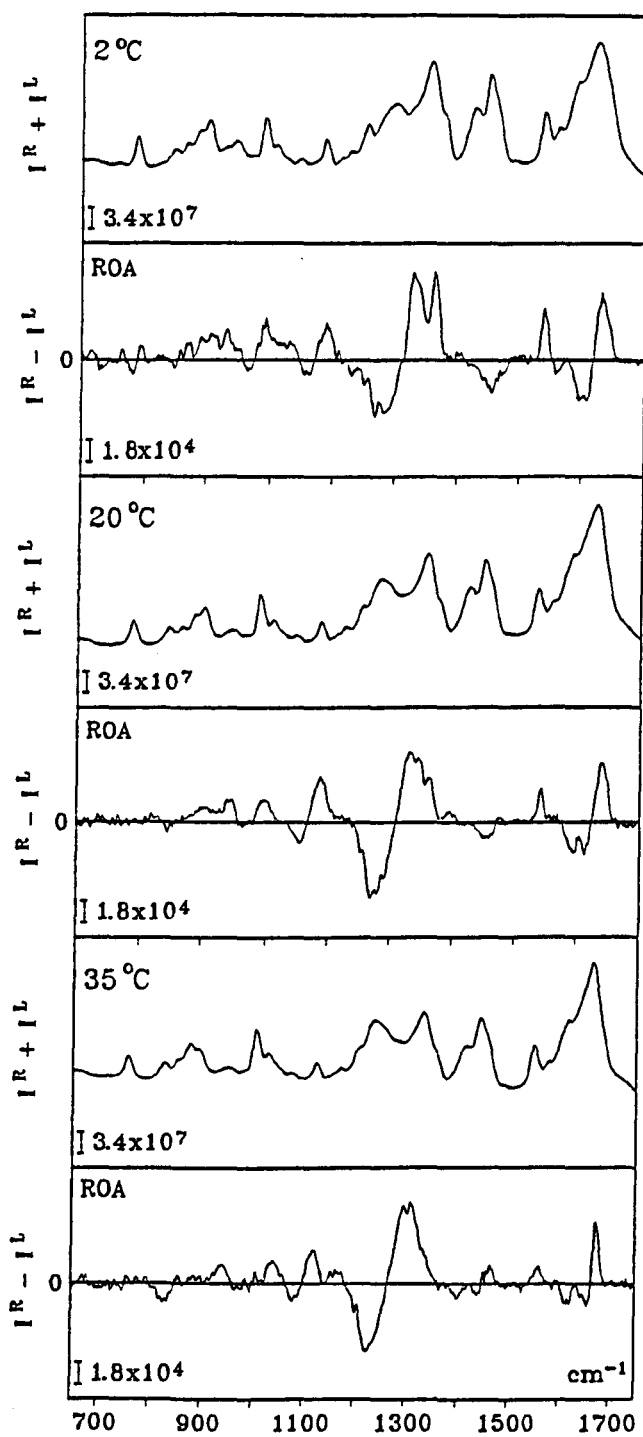


Figure 7.3 Backscattering Raman and ROA of molten globule  $\alpha$ -lactalbumin at 2, 20 and 35°C

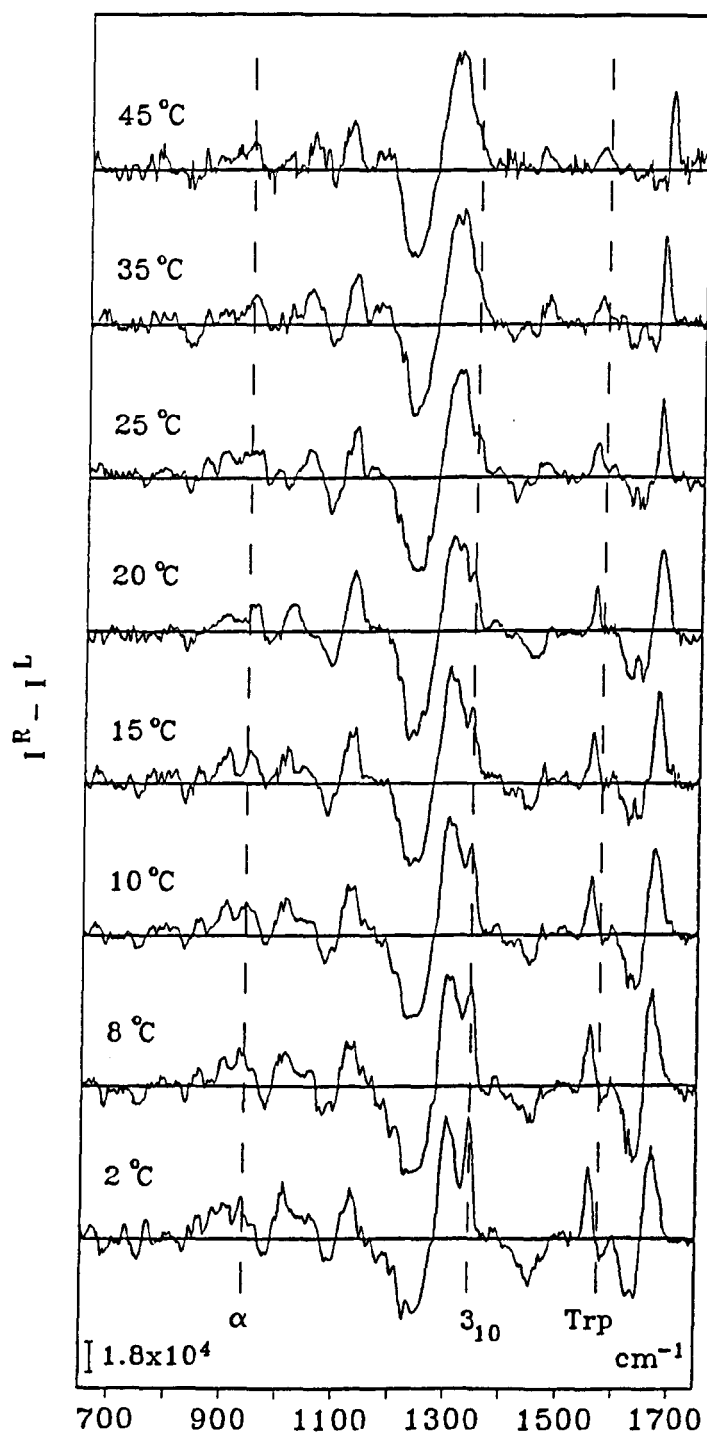


Figure 7.4 Backscattering ROA of molten globule  $\alpha$ -lactalbumin at 45 - 2°C



From an inspection of Figure 7.3 and Figure 7.4 there are two trends that can be observed as the temperature is changed. At 35°C some ROA secondary structure bands which can still be seen in the backbone skeletal stretch region where there is a resemblance to the native protein but there has been a loss of intensity and changes of detail in parts of the ROA band structure. As the temperature is lowered the amount of secondary structure gradually increases. The second trend that can be observed from the ROA spectra is that the loop structure in the extended amide III region changes rapidly (Wilson *et al.*, 1995; Wilson *et al.*, 1996). The sharp positive ROA band at 1340 cm<sup>-1</sup> is an ideal band to monitor this change in the tertiary fold. It is also noticeable that at 2°C the molten globule ROA is very similar to that of the native protein. This suggests that the complete structure (including the tertiary fold) is native-like yet the sample retains the molten globule status (since it still exhibits the loss of near-UV CD and the fluorescence changes).

The ROA spectrum at 35°C is consistent with the standard ideas of molten globule structure. In the backbone skeletal stretch region it can be observed that some of the secondary structure has remained and retains a similarity to the native protein, yet there is a reduction in the intensity and structure of the ROA bands. The extended amide III region at 35°C is dominated by a conservative couplet with the negative component peaking at 1230 cm<sup>-1</sup> and the positive part which is slightly broader at 1300 cm<sup>-1</sup>. The shape of this couplet is very similar to that shown by the unfolded lysozyme (Ford *et al.*, 1995) which suggests that most of the tertiary fold is lost. On going to 45°C this couplet becomes even more like that of unfolded lysozyme. This implies that at these high temperatures the loop structure must have a similar conformational heterogeneity and mobility to that found in an unfolded protein. At 35°C the amide I ROA couplet becomes much sharper and similar to the situation in unfolded lysozyme, with the couplet shifted by ~8 cm<sup>-1</sup> to a higher wavenumber. This is another indication that the conformational heterogeneity has increased at this high temperature.

Fluorescence experiments ( $\lambda_{\text{exc}} = 295 \text{ nm}$ ) on the acid molten globule sample at 35°C showed a red shift of the maximum from 334 nm in the native

protein at pH 8.0 to 346 nm with a small increase in intensity (e.g. Lala & Kaul, 1992). A comparable result was obtained in the near-UV CD since at 35°C the negative intensity that is observed in the native protein is lost in the molten globule. These described changes in fluorescence and near-UV CD experiments are usually related to changes in the environments and conformations of some of the aromatic residues. The red shift in the fluorescence usually arises from an increased exposure to solvent water which will provide a more polar environment which induces the red shift. The increase of the intensity implies that the tertiary interactions that have the ability to quench the fluorescence in the native state have been lost (Hayer-Hartl *et al.*, 1994). The loss of intensity in the near-UV CD is due to a time-averaged symmetrization of the environments of the aromatic side-groups which is a consequence of the fluctuating tertiary structure (Dolgikh *et al.*, 1981). However, using the near-UV CD to predict the absence or the presence of the native tertiary fold does not now seem to be definitive because the absence of near-UV CD does not necessarily mean that the tertiary structure is unfolded (Ku wajima, 1989; Ptitsyn, 1992; Okasaki *et al.*, 1994; Sugai & Ikeguchi, 1994; Ewbank *et al.*, 1995; Fink, 1995). The near-UV CD results can therefore be misleading and it is better to assume that the presence of near-UV CD from aromatic residues just gives a signal that is representative of the packing of the tertiary fold (Christensen & Pain, 1994), rather than saying that the loss of near-UV CD definitely means that the tertiary fold is lost. The actual mechanism by which near-UV CD is generated in aromatic chromophores is actually not yet fully understood (Woody, 1994) and it remains unclear as to why near-UV CD shows nothing for low temperature acid molten globule  $\alpha$ -lactalbumin where ROA has shown that the tertiary fold is highly native-like with the tryptophans rigid, which will now be discussed (this might be due to the solvated tryptophans being held rigidly to the backbone).

Considering now the 2°C ROA spectrum in Figure 7.3 it is apparent that the ROA spectrum is very similar to that of native  $\alpha$ -lactalbumin as shown in Figure 6.3. In the backbone skeletal stretch region the ROA of the 2°C ROA spectrum is very similar to that of the native protein, even the detailed structure has reappeared. As

opposed to the sharp amide I couplet in the 35°C ROA spectrum, the shape and position of the amide I is also native-like. The  $\alpha$ -helix amide III couplet which has its negative part at 1265  $\text{cm}^{-1}$  and positive component at 1296  $\text{cm}^{-1}$  is also suggesting a native-like structure. The loop signatures that are not present at high temperatures occur at 2°C which suggests that not only is the secondary structure native-like but so also is the tertiary fold. The negative ROA bands at 1223 and 1243  $\text{cm}^{-1}$  and the positive band at 1340  $\text{cm}^{-1}$  are indicative of the native  $\alpha$ -lactalbumin tertiary fold. This sample at 2°C is still classed as a molten globule since it still shows the red shift in the fluorescence and the loss of intensity in the near-UV CD as did the spectra at 35°C, which indicates that the environments of the aromatic residues remain the same at the high and low temperatures. However, the amide III region is not as sharp as that exhibited by the native protein and there is a slight broadening and increase in noise of the couplet.

The appearance of the 20°C ROA spectrum of the acid molten globule of  $\alpha$ -lactalbumin (middle of Figure 7.3) is intermediate between the two extremes and shall now be discussed. The backbone skeletal stretch and amide I region still look highly native-like, indicating that most of the secondary structure is still intact. On the other hand, the amide III is quite different from either the 2°C or 35°C spectra. As expected the  $\alpha$ -helix couplet positive at 1296  $\text{cm}^{-1}$  and negative at 1265  $\text{cm}^{-1}$  is still present but the positive ROA band at 1340  $\text{cm}^{-1}$  has lost much of its intensity. The two negative ROA bands at 1223 and 1243  $\text{cm}^{-1}$  which are present in the native protein and the 2°C molten globule form are replaced by a single sharp negative bend at 1230  $\text{cm}^{-1}$  as found in the 35°C spectrum. The acid molten globule state of  $\alpha$ -lactalbumin at 20°C therefore contains a secondary structure similar to that of the 2°C molten globule but has lost much of the tertiary loop structure and hence much of the tertiary fold.

Due to some of the ROA bands changing sufficiently over the temperature range it is possible to make a plot of ROA intensity against temperature for acid molten globule  $\alpha$ -lactalbumin. The following ROA bands were selected for the plot: the positive band at 931  $\text{cm}^{-1}$  which is an  $\alpha$ -helix signature; the positive ROA band at

1340  $\text{cm}^{-1}$  assigned to  $3_{10}$ -helix loop; and we used the peak-to-peak amplitude of the tryptophan couplet associated with Raman bands at 1555 and 1580  $\text{cm}^{-1}$ . The plots are shown in Figure 7.5.

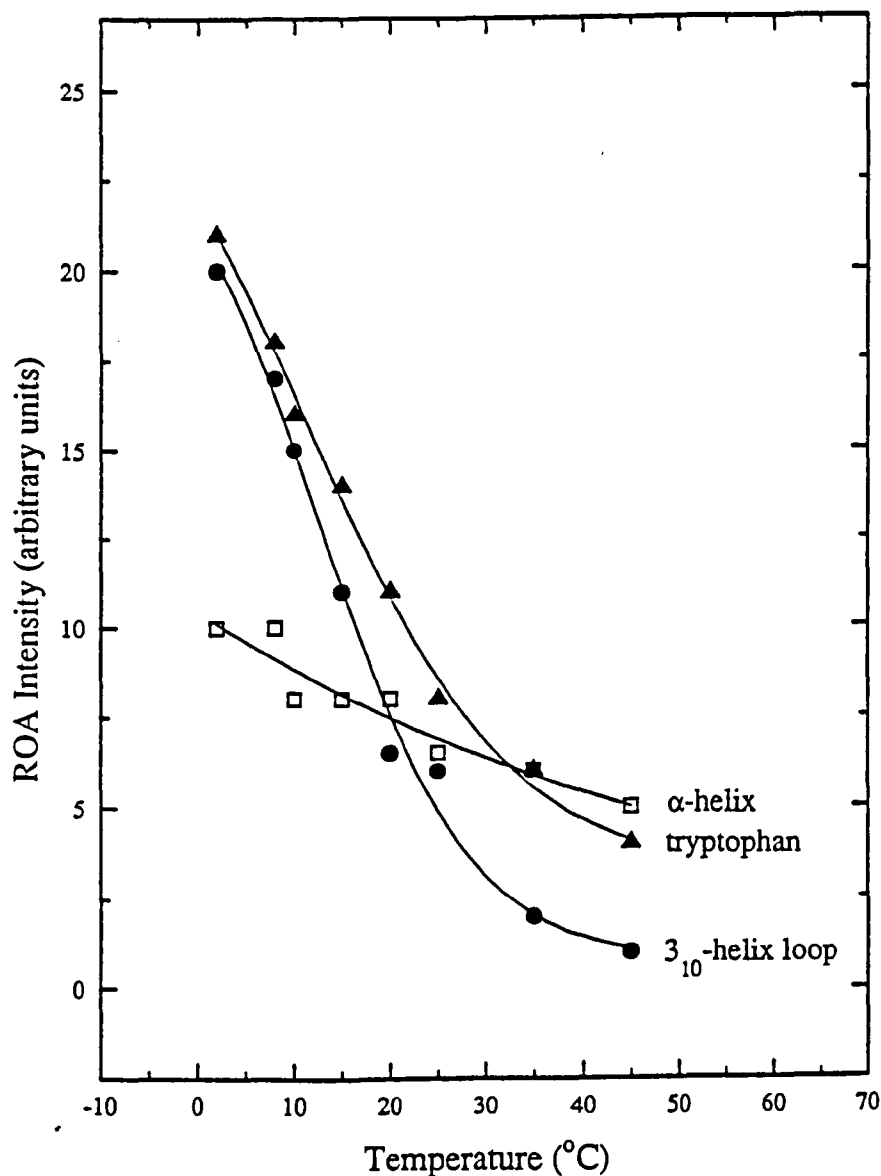


Figure 7.5 Plot of ROA intensity against temperature for acid molten globule  $\alpha$ -lactalbumin

From an inspection of Figure 7.5 it is clear that the  $\alpha$ -helix undergoes a gradual change over the 2 - 45°C temperature range which is characteristic of a non-cooperative process. In contrast, the  $3_{10}$ -helix loop and tryptophan plots are sigmoidal, which is characteristic of a cooperative process (Dill & Shortle, 1991; Chan *et al.*, 1995).

The  $\alpha$ -helix ROA results agree with far-UV CD experiments which also indicated that when the temperature increases the amount of  $\alpha$ -helix decreases gradually via a non-cooperative process (Kuwajima, 1989). A major advantage that ROA has over UV CD becomes apparent here because UV CD cannot detect the presence of loop structure and so cannot monitor the changes in loop structure and hence the tertiary fold, as the temperature is varied. The near-UV CD is also not helpful due to the aromatic residues being fully exposed to solvent in the molten globule at all temperatures. This situation arises due to the different mechanisms in ROA and in near-UV CD from which the effects are generated. ROA probes the local geometry of the aromatic residues relative to the peptide backbone *via* the coupling of vibrational coordinates in the aromatic residues with those of the connecting covalent bonds.

The cooperative nature of the changes in  $3_{10}$ -helix loop structure with temperature has far reaching implications for changes of the complete tertiary fold. From the crystal structure of  $\alpha$ -lactalbumin (Acharya *et al.*, 1989) the  $3_{10}$ -helix is distributed all over the structure; in particular residues 12 - 21, 76 - 82, 101 - 104 and 115 - 119. Any change in  $3_{10}$ -helix will therefore affect the tertiary fold of the whole protein. In the extended amide III region the other loop signatures can be seen to follow a similar pattern as the 1340  $\text{cm}^{-1}$  band, although it would not be as easy to extract the required data due to complications arising with other overlapping bands. When the acid molten globule  $\alpha$ -lactalbumin was tested under a similar set of conditions using differential scanning calorimetry (DSC) no latent heat and therefore no transition was detected. This result is shown in Figure 7.6 and agrees with other experimental data (Griko *et al.*, 1994a).

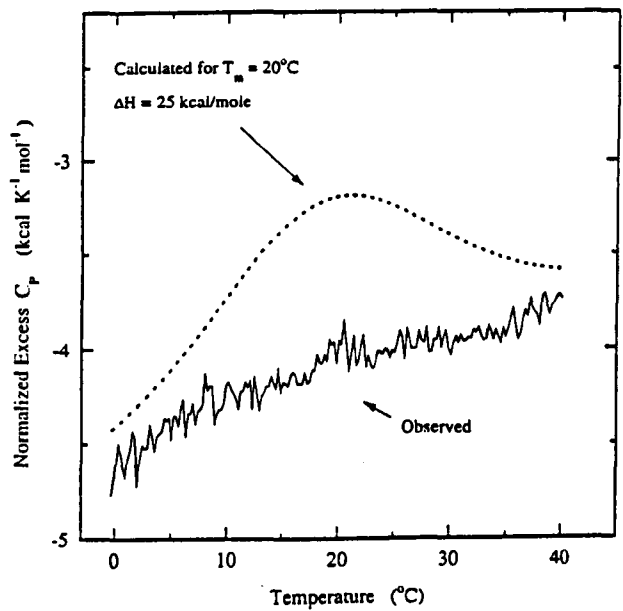


Figure 7.6 DSC of acid molten globule  $\alpha$ -lactalbumin

A similar situation was also found for the fluorescence of the acid molten globule  $\alpha$ -lactalbumin, as shown in Figure 7.7, where no transition was observed.

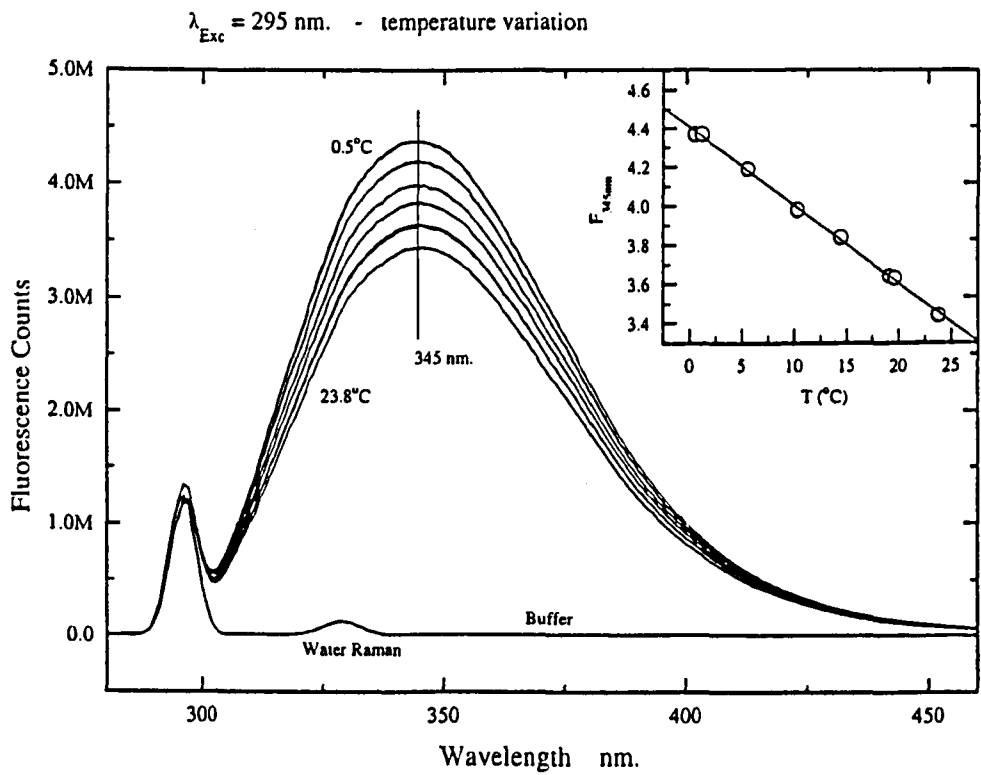


Figure 7.7 Fluorescence of acid molten globule  $\alpha$ -lactalbumin

If the transition was first order and if the points in the  $3_{10}$ -helix loop are used in a conventional van't Hoff analysis it can be calculated that an enthalpy change of  $\Delta H \approx 25 \text{ k cal mol}^{-1}$ , with a corresponding transition temperature of  $\sim 20^\circ\text{C}$  should occur. However, DSC observes no latent heat which implies that the transition cannot be first order and that a van't Hoff analysis is inappropriate. Cooperativity but no latent heat are the characteristics of a second-order phase transition (Landau & Lifshitz, 1980). A second-order phase transition can be described as having no discontinuity in the properties and thermodynamic functions of the system at the transition temperature although there can be discontinuities in the rates of change. In condensed matter physics anything apart from a first order transition is now called a *continuous transition* to avoid problems with the original Ehrenfest classification and this nomenclature will be used here (Binney *et al.*, 1992). Another reason for not calling the transition second-order is that the characteristic discontinuity is not present in the latent heat since no step is perceivable in the DSC trace in Figure 7.6. There is also the possibility that this observed transition in molten globule  $\alpha$ -lactalbumin could be a spin-glass transition (Austin & Chen, 1992).

In a first-order transition different phases will co-exist at the transition point and this will result in an abrupt structural change when the transition occurs. This is in contrast to a continuous transition where the structural differences between the different phases disappear at the critical point (Landau & Lifshitz, 1980; Binney *et al.*, 1992). All transitions require an *order parameter* which is an observable quantity that represents the degree of ordering of the microscopic system and which varies continuously through non-zero values below the initial temperature to zero above it. A phase transition that contains these qualities is termed an *order-disorder* transition and is always accompanied by a reduction of symmetry in the low-temperature phase. These ideas tie in with the loop signatures in the molten globule and especially the positive ROA band at  $1340 \text{ cm}^{-1}$  which originates in  $3_{10}$ -helix loop structure, since these loop signatures could be classed as an order parameter that is related to the tertiary fold. This order parameter would be a pseudoscalar (i.e. a

number with no directional properties but which changes sign under space inversion) and therefore the loss of the signal in the disordered phase is monitoring a time-averaged loss of the chirality in the system.

The tryptophan ROA couplet associated with the 1555 and 1580  $\text{cm}^{-1}$  Raman bands also follows a sigmoidal curve similar to that of the  $3_{10}$ -helix loop but with the transition range shifted to a slightly higher temperature. There is also a small aromatic couplet centred at 1615  $\text{cm}^{-1}$  which is also due to tryptophan (or other aromatic) vibrations but is too small to analyse its variation with temperature. The larger couplet is very similar at 2°C and 20°C, but virtually collapsed at 35°C and 45°C. The ROA is therefore suggesting that at 2°C and 20°C the tryptophan residues in the acid molten globule have a rigidity similar to that found in the native protein. Independent evidence from fluorescence polarization measurements showed by analysing the rotational relaxation times of the tryptophans in the molten globule that they were similar to the native state and perhaps even more rigid (Dolgikh *et al.*, 1981). By monitoring the tryptophans it is possible to describe the state of the tertiary fold because three of the four tryptophans (60, 104 and 118) occur in the loop regions and will therefore become more mobile as the tertiary fold is lost. The fourth tryptophan which occurs at residue 26 is situated in an  $\alpha$ -helix region and will be relatively unaffected under the described experimental conditions and could perhaps explain why the tryptophan sigmoidal curve is shifted relative to the  $3_{10}$ -helix loop sigmoidal curve.

The temperature-dependent ROA spectra are also able to provide information about the aliphatic side-groups. There appears to be large changes in the aliphatic side-group signatures in the range ~1400 - 1520  $\text{cm}^{-1}$ . Going from 2°C to 20°C the features have collapsed into a small negative feature at the low-wavenumber end and all of the weak positive ROA at the high wavenumber end has disappeared. The ROA from the aliphatic side-groups is generated by coupling between  $\text{CH}_2$  and  $\text{CH}_3$  side-group deformations and vibrational coordinates of the backbone amide II mode and aromatic residues. When this coupling mechanism is quenched it suggests that the aliphatic side-groups are much more mobile. An



interesting conclusion that can be inferred from this is that in the molten globule there is sufficient space for the free movement of small symmetric aliphatic groups but not enough room for the larger and less symmetric aromatic groups to have a high degree of mobility (Ptitsyn, 1992). The coupling of the amide II to the vibrational coordinates of the side-groups has been found to be strongly dependent on hydrogen bonding (Krimm & Bandekar, 1986) which follows from the idea that the rate-limiting step in protein unfolding is the cooperative breaking of the hydrogen bond network (Kiefhaber *et al.*, 1995).

The  $\beta$ -sheet region also shows a change on going from 20°C to 25°C, since the positive  $\beta$ -sheet band shifts from 1006  $\text{cm}^{-1}$  to 1030  $\text{cm}^{-1}$ .

$\alpha$ -lactalbumin is also known to form a molten globule when the  $\text{Ca}^{2+}$  has been removed from the binding loop and although this situation will be discussed in detail in the next section it is relevant to introduce some recent experimental results for this type of molten globule (Wu *et al.*, 1995). Wu *et al.* recently studied two variants of  $\alpha$ -lactalbumin that are molten globules in the absence of  $\text{Ca}^{2+}$  and deduced that the  $\alpha$ -helical domain of  $\alpha$ -lactalbumin exhibited native-like molten globule properties while the  $\beta$ -sheet domain which contains the calcium binding loop is almost completely unstructured. They used UV CD to determine the presence of the molten globule. This is perhaps an over-simplistic view of the situation though since the ROA results directly show that the presence of the tertiary fold is very dependent on the temperature. The acid molten globule form at 20°C could well be classed as a bipartite structure but, due to the close resemblance of the 2°C ROA spectrum to that of the native protein implying the existence of a native-like backbone, the 2°C acid molten globule is definitely not consistent with the bipartite structural theory.

A recent theory (Shakhnovich & Finkelstein, 1989a,b) concluded that the primary step in protein denaturation is the cooperative loss of the tight packing of the side-chains that stabilise the tertiary structure (Ptitsyn, 1992), with or without loss of the native-like tertiary fold. Ptitsyn has suggested the use of "native-like molten globule" and "disordered molten globule" to distinguish if the structure has a native-

like tertiary fold. Using this terminology and from the fluorescence and near-UV CD results all of the samples shown in Figure 7.4 would have the molten globule status. The lower temperature samples such as the 2°C sample would be termed native-like molten globules, while the 35°C and 45°C samples would be classed as disordered molten globules. The 20°C sample because it retains some of the tertiary fold could be classed as a partially disordered molten globule.

ROA has therefore shown that the secondary structure remains mostly intact in the acid molten globule state over the 2 - 35°C temperature range and decreases gradually as the temperature is raised. The tertiary fold on the other hand changes dramatically over the described temperature range, at 2°C the tertiary fold is almost complete and native-like whereas at 35°C it is almost completely disordered. The amount of secondary structure appeared to vary *via* a gradual non-cooperative process while the tertiary fold was dependent on a cooperative transition. The rigid loop structure in the tertiary fold therefore interconverts with disordered mobile structure in the loops that link the regions of rigid secondary structure. This transition is a continuous, order-disorder phase transition which could be of vital importance in explaining how proteins find their native-like tertiary folds. Although the initial interactions are short-range the continuous nature of the transition means that long-range interactions become effective in the critical region which becomes the dominating factor in the way that a system chooses the way that it evolves (Binney *et al.*, 1992).

These ROA results agree with protein folding models that depend on molten globule intermediates (Christensen & Pain, 1994) in which the development of the tertiary fold depends on a significant amount of pre-existing native-like secondary structure. The important extra ingredient that ROA adds is that the development of the tertiary fold is dependent on a continuous phase transition.

### Calcium-Free $\alpha$ -Lactalbumin

When the  $\text{Ca}^{2+}$  is removed from the binding loop in  $\alpha$ -lactalbumin and the sample dissolved in a low salt solution at neutral pH the resultant sample is said to be a molten globule (Kuwajima, 1989). After the acid molten globule study it was therefore natural to study the calcium-free (apo)  $\alpha$ -lactalbumin form over a range of temperatures. As described in the experimental section the samples were dissolved in Tris buffer at pH 8.0. The backscattered Raman and ROA spectra of apo  $\alpha$ -lactalbumin at 2°C, 25°C and 60°C are shown in Figure 7.9. In Figure 7.8 the ROA spectrum of the calcium-bound (holo) native  $\alpha$ -lactalbumin in Tris buffer at pH 5.4 and pH 8.0 are displayed so that we have a direct comparison (the apo  $\alpha$ -lactalbumin was not sufficiently soluble at a lower pH). It can be observed that the  $\alpha$ -lactalbumin ROA spectra at these different pH's are almost identical.

The 2°C apo  $\alpha$ -lactalbumin ROA spectrum shown in the top third of Figure 7.9 is very similar to that of the holo form at pH 8.0 in Figure 7.8. The secondary structure of the apo form at 2°C is virtually unchanged compared to the holo form but perhaps of greater importance is that the tertiary structure appears to be very similar in both spectra. It is clear that the two negative ROA bands at 1223 and 1243  $\text{cm}^{-1}$ , and the positive band at 1340  $\text{cm}^{-1}$ , which are loop signatures, appear with very similar shape and intensity in both spectra. This indicates that the apo  $\alpha$ -lactalbumin at 2°C has a completely native-like tertiary fold. The  $\text{Ca}^{2+}$  is bound between residues 74 to 90 and these results imply that the double-loop region remains rigid. It is not classed as a molten globule since the fluorescence and near-UV CD results were the same as for holo  $\alpha$ -lactalbumin; namely a fluorescence peak at 334 nm and a negative near-UV CD intensity, which shows that apo  $\alpha$ -lactalbumin is not a molten globule at 2°C.

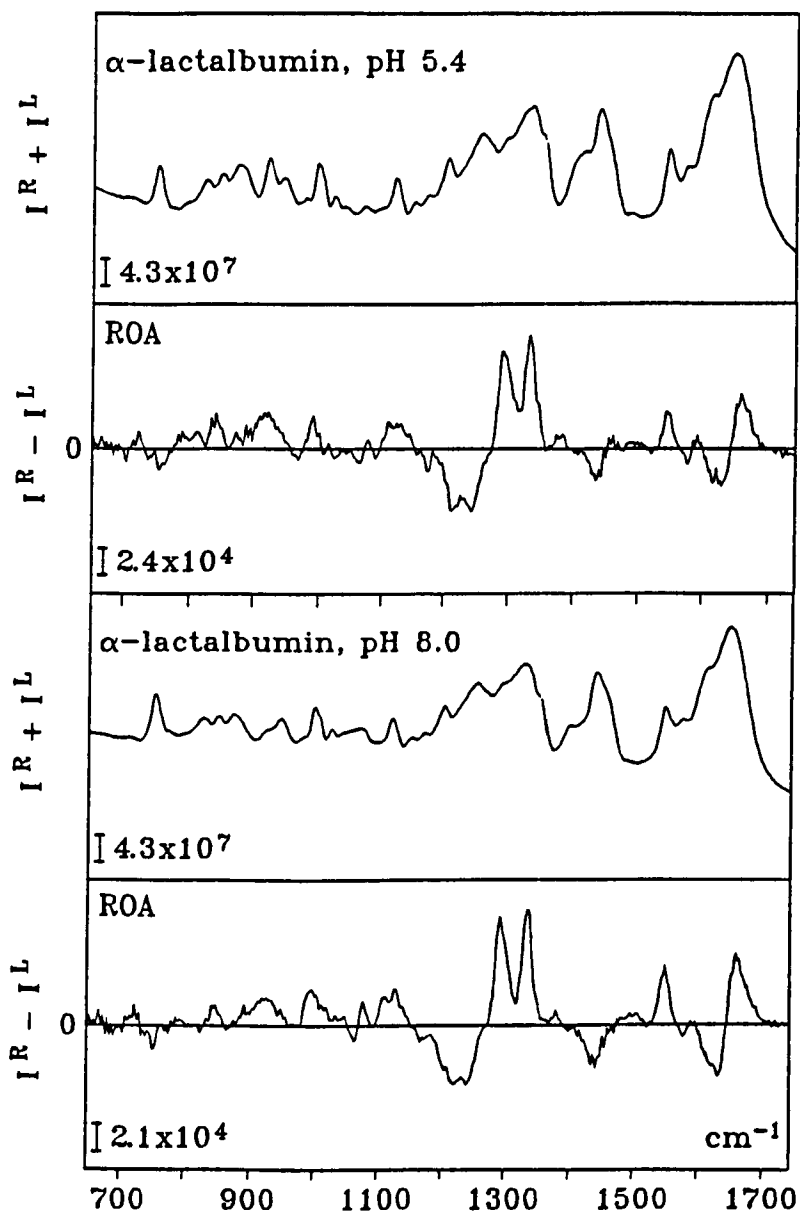


Figure 7.8 Backscattered Raman and ROA of  $\alpha$ -lactalbumin at pH 5.4 and pH 8.0

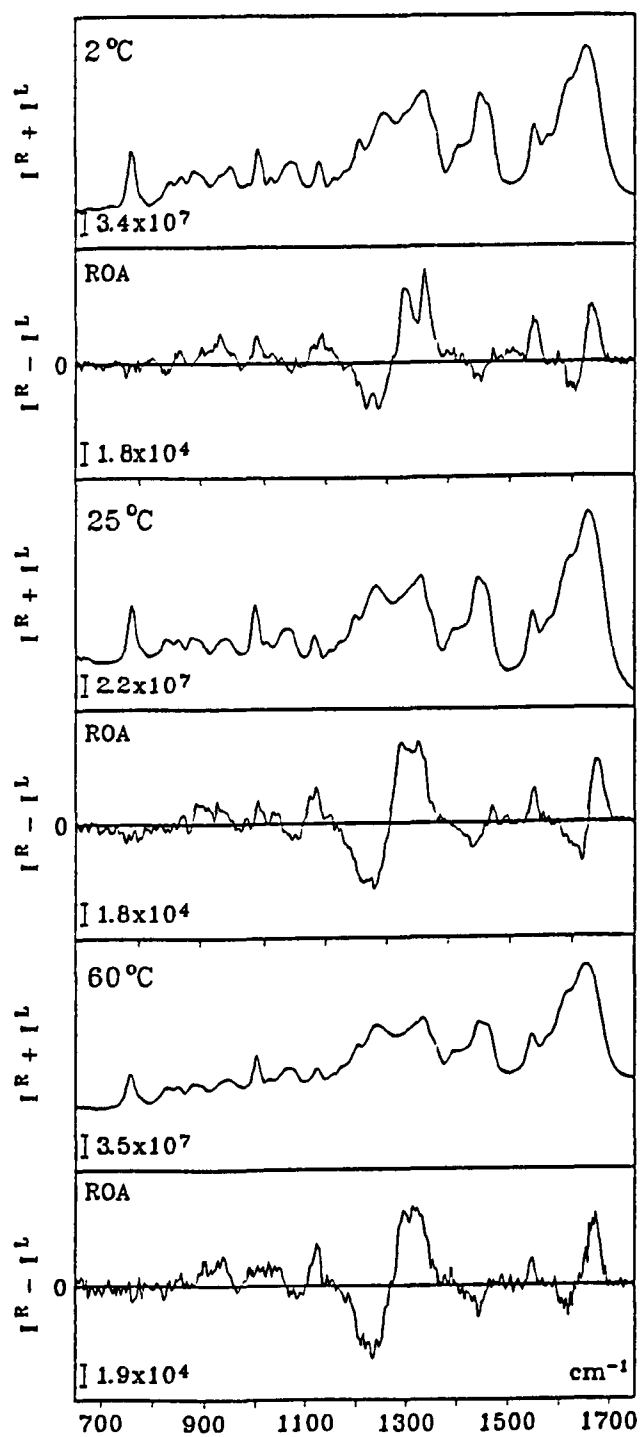


Figure 7.9 Backscattered Raman and ROA of apo  $\alpha$ -lactalbumin at 2, 25 and 60°C

In the middle of Figure 7.9 the apo  $\alpha$ -lactalbumin at 25°C is displayed. This spectrum is obviously different from the 2°C spectrum, with the loop signatures undergoing change and there is an increased level of noise in the spectrum. The negative ROA bands at 1223 and 1243  $\text{cm}^{-1}$  are still present but they have lost their sharpness and are now much broader. The positive ROA band at 1340  $\text{cm}^{-1}$  has reduced in intensity which indicates that there is a reduction in the amount of  $3_{10}$ -helix loop structure. This change may occur in residues 76 to 82 where the calcium is bound and may well be indicating that on increasing the temperature from 2°C to 25°C the apo  $\alpha$ -lactalbumin has undergone a transition from a native-like state to a molten globule one. There is also an interesting comparison with lysozyme since on going from the crystal to aqueous solution the  $3_{10}$ -helix found in residues 79 to 84 loses its well-defined structure.

DSC shows the native to molten globule transition occurring at  $\sim 20^\circ\text{C}$  as shown in Figure 7.10

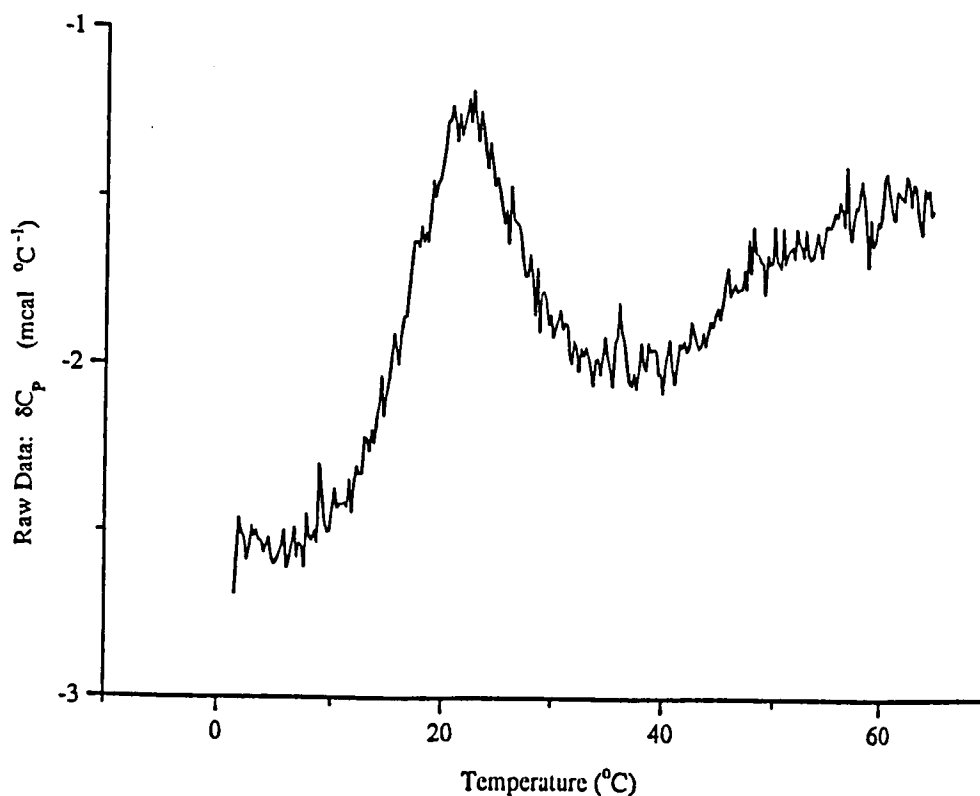


Figure 7.10 DSC of apo  $\alpha$ -lactalbumin

Fluorescence measurements (Permyakov *et al.*, 1981; Kronman *et al.*, 1981; Murukami *et al.*, 1982; Ewbank & Creighton, 1993) on apo  $\alpha$ -lactalbumin showed that on raising the temperature the observed intensity slowly increased and shifted towards the red and stabilized at a maximum of 346 nm at just below 25°C which is what is expected for the molten globule. The temperature-dependent fluorescence spectra were performed and are shown in Figure 7.11.

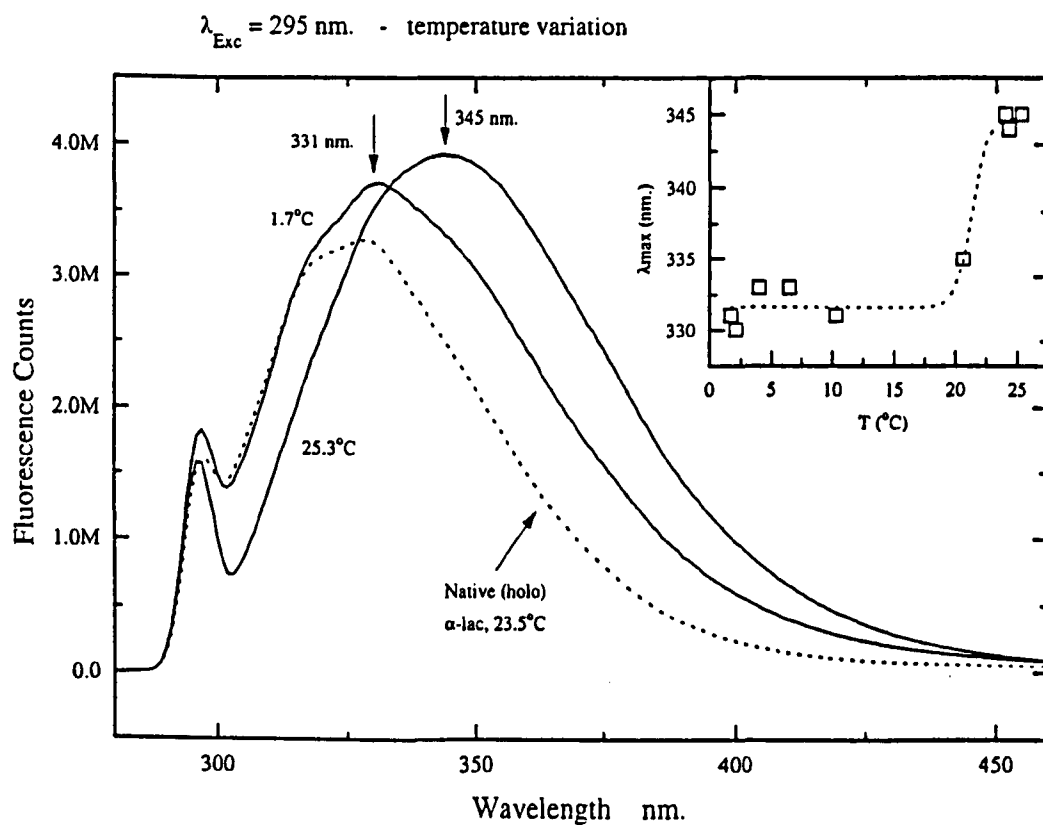


Figure 7.11 Fluorescence of apo  $\alpha$ -lactalbumin

As the temperature is raised to 35°C although the loop signatures continue to degrade they are still present as shown in Figure 7.12 which shows five ROA spectra of apo  $\alpha$ -lactalbumin at different temperatures. This is in direct contrast to the acid molten globule at 35°C, which has almost completely lost its tertiary fold and is very similar to the ROA of unfolded lysozyme in the extended amide III region.

There is a cooperative transition at ~20°C to the molten globule state. Above 25°C the apo  $\alpha$ -lactalbumin is in a native-like molten globule state with the tertiary fold relatively intact and a significant amount of secondary structure. As the temperature is increased to 35°C and 60°C, the secondary structure and tertiary fold are not drastically affected. This is opposed to the situation in the acid molten globule case where the secondary structure and tertiary fold had all but collapsed at 50°C. ROA has therefore shown that apo molten globule  $\alpha$ -lactalbumin is much more stable than the acid form.



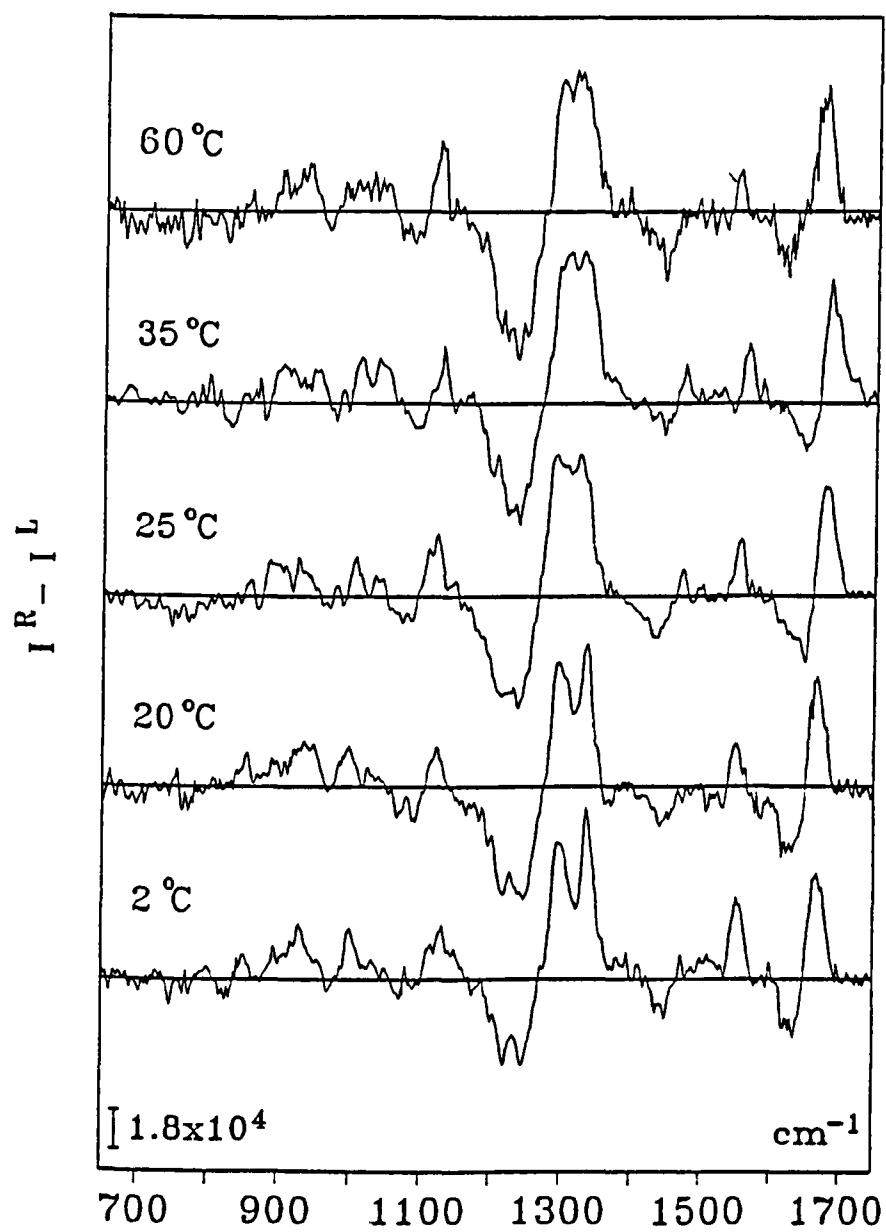


Figure 7.12 Backscattered ROA of apo  $\alpha$ -lactalbumin at 60, 35, 25, 20 and 2°C

Attempts to characterise apo  $\alpha$ -lactalbumin has led to much discussion (Ewbank *et al.*, 1995; Okazaki *et al.*, 1995) and in some cases conflicting conclusions. GroEL binds to the molten globule state of  $\alpha$ -lactalbumin (Kuwajima, 1989) and may therefore be used to detect the presence of a molten globule state. On removal of  $\text{Ca}^{2+}$  Okazaki *et al.* (1994) observed no signal in the near-UV CD spectra and therefore assumed that the tertiary structure was unfolded; but were then surprised that GroEL had not bound to what they thought was molten globule  $\alpha$ -lactalbumin. A similar set of experiments were carried out by Hayer-Hartl *et al.* (1994) who also found that GroEL did not bind to their apo  $\alpha$ -lactalbumin but due to their hypothesis that an equilibrium was in effect between the native and molten globule conformations were not in the least surprised. The actual binding constant of GroEL will depend on the salt concentration and will be restricted from binding in the apo  $\alpha$ -lactalbumin due to the four native disulfide bonds. On reducing the four disulfides the GroEL was found to bind strongly.

The GroEL binding experiments were performed at 25°C and the ROA results indicate a native-like molten globule at this temperature. At 25°C apo  $\alpha$ -lactalbumin therefore has a native-like tertiary fold and it is probably this that prevents GroEL from binding, which agrees with the recent electrospray ionization mass spectrometry results of Robinson *et al.* (1994) who deduced that only a disordered molten globule would bind to GroEL and apo  $\alpha$ -lactalbumin does not bind since it exists in a more ordered state at 25°C than the acid form and is therefore in agreement with ROA.

### **Comparison of ROA Studies of Molten Globule $\alpha$ -Lactalbumin with other Spectroscopic Techniques**

In spite of the fact that NMR cannot provide a complete solution structure of any form of apo or holo  $\alpha$ -lactalbumin (due to aggregation and dynamic fluctuations) it has provided information that is in agreement with the recent ROA studies

(Alexandrescu *et al.*, 1993; Chyan *et al.*, 1993). The NMR studies which were performed above 35°C revealed that acid molten globule  $\alpha$ -lactalbumin retains the majority of its secondary structure, particularly in the B and C  $\alpha$ -helical region. NMR also revealed the lack of any native-like tertiary loop structure which agrees with ROA because at 35°C and especially at 45°C the ROA loop bands, and hence the tertiary fold, was almost completely gone. The NMR results on the acid molten globule above 35°C also revealed that the aromatic residues which in the native protein make up part of the hydrophobic core of the helical domain in the native protein are unordered. What was found though was that there was a new smaller aromatic cluster (Alexandrescu *et al.*, 1993) which agrees with the ROA at 35°C and 45°C because at these temperatures the aromatic ROA bands although much weaker are still present. If NMR measurements were to be performed at 20°C perhaps more of a native-like clustering might be observed for the acid molten globule  $\alpha$ -lactalbumin. Urea induced unfolding (Shimizu *et al.*, 1993) NMR studies also showed that the unfolding transitions as monitored by changes in the individual aromatic resonances did not coincide, implying that the unfolding of the acid molten globule state of  $\alpha$ -lactalbumin is not a cooperative two-state process. This conclusion agrees with ROA results which found that the secondary structure changed gradually over temperature in the acid molten globule state while the loop structure changed cooperatively. NMR also concluded that the apo  $\alpha$ -lactalbumin has a similar three dimensional conformation to the native form (Berliner *et al.*, 1987), although this is not based on complete 3-dimensional structures. This agrees with the ROA data which showed that apo  $\alpha$ -lactalbumin had a native-like tertiary fold even at above room temperatures. The loss of  $\text{Ca}^{2+}$  does not lead to any significant change in the peptide backbone apart from perhaps the loss of some of the  $3_{10}$ -helix in the calcium binding loop. It is also reassuring that a recent VCD study also came to the same conclusions (Keiderling *et al.*, 1994).

The apo form of  $\alpha$ -lactalbumin has also been studied (in  $\text{D}_2\text{O}$ ) *via* FTIR (Prestrelski *et al.*, 1991b). From an examination of the amide I band structure it was concluded that on the removal of  $\text{Ca}^{2+}$ , the  $\text{Ca}^{2+}$  binding loop was found to undergo

local unfolding. These FTIR results which were carried out at 25°C concur with ROA results which indicated the possible loss of the  $3_{10}$ -helix within the binding loop above 25°C.

### Ligand Binding: Lysozyme with Tri-N-Acetyl Glucosamine

The backscattered Raman and ROA spectra of native lysozyme and lysozyme bound to the saccharide inhibitor NAG<sub>3</sub> are shown in Figure 7.13. It is interesting to observe that the ROA spectra of the unbound and ligand bound form of the lysozyme are different. ROA is therefore able to monitor subtle conformational changes in the protein which could be useful because enzyme catalysis can involve similar changes. The ROA of NAG<sub>3</sub> itself was also obtained (not shown) and it was found to have a weak ROA spectrum indicating that the changes in the ligand bound lysozyme ROA spectrum are due to conformational changes of the protein and are not intrinsic to NAG<sub>3</sub> itself. Lysozyme is an excellent example for this study because not only is the crystal structure known for the native lysozyme (McKenzie & White, 1991) but that for the NAG<sub>3</sub> bound form is also known (Blake *et al.*, 1967). In the crystal structure of the ligand bound form small changes in the lysozyme structure were indeed observed.

The main difference in the ROA spectra on binding the NAG<sub>3</sub> is the large increase in the ROA band at 1340 cm<sup>-1</sup> and a decrease in the underlying amide III couplet. The increase in the 1340 cm<sup>-1</sup> band implies that there is an increase in rigid loop structure, probably that of 3<sub>10</sub>-helix. This idea fits with the decrease of the underlying amide III couplet which from the study of the unfolded lysozyme we know is due to mobile structure.

These ROA observations are consistent with X-ray (Blake *et al.*, 1981) and NMR (Bernard *et al.*, 1990; Kumagai *et al.*, 1993; Lamb *et al.*, 1994) data where in the crystal the residues from 72 - 76 were observed to become more rigid and in solution the NMR showed that on binding the inhibitor it affected the loop residues from 70 - 75.

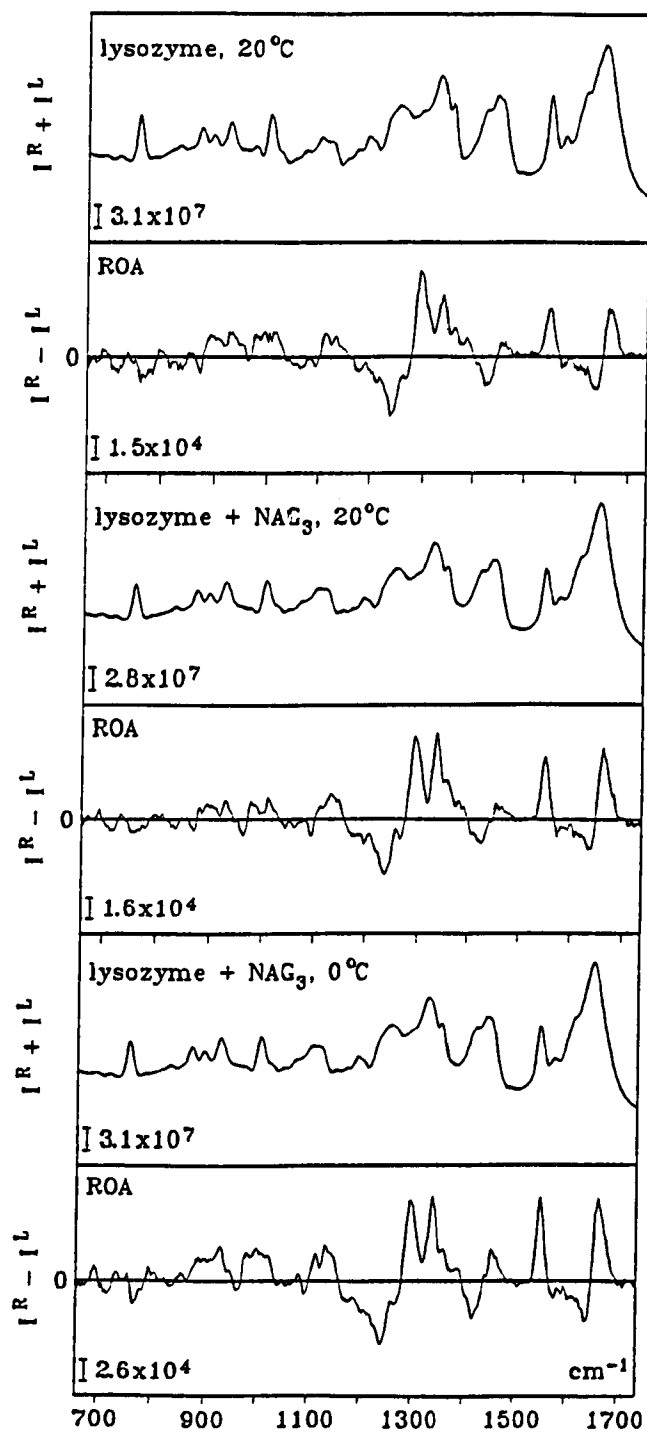


Figure 7.13 Backscattered Raman and ROA of lysozyme and NAG<sub>3</sub> bound lysozyme at 20 and 0°C

There has also recently been a conventional Raman study where the monomer NAG was bound to lysozyme (Bertoluzza *et al.*, 1992). They found an increase in rigidity of the lysozyme with a 15% increase in  $\alpha$ -helix accompanied by a corresponding decrease in random coil structure. This agrees with the ROA study in the aspect of increasing rigidity in the protein but conflicts with the type of structure that is increasing in rigidity. It is however not a direct comparison since in ROA studies the trimer  $\text{NAG}_3$  is used whereas in the conventional Raman study the monomer of NAG was bound (they perhaps see  $3_{10}$ -helix as  $\alpha$ -helix).

The ROA of the lysozyme bound to the  $\text{NAG}_3$  was also obtained at 2°C. It is intriguing that the positive tryptophan ROA band at  $1555\text{ cm}^{-1}$  and the backbone skeletal stretch regions increase in intensity. This suggests that at low temperatures the saccharide bound lysozyme is becoming even more rigid (see next section).

### Temperature-Dependence of Native Lysozyme ROA

Surprising results were obtained from studies of the temperature-dependence of the ROA spectrum of native lysozyme. Significant changes in the spectra were monitored over the 2 - 50°C temperature range.

In Figure 7.14 and Figure 7.15 the backscattered Raman and ROA spectra are displayed at six different temperatures within the described temperature range, namely 2°C, 8°C, 12°C, 15°C, 20°C and 50°C. The trends observed in the temperature-dependent spectra will now be discussed.

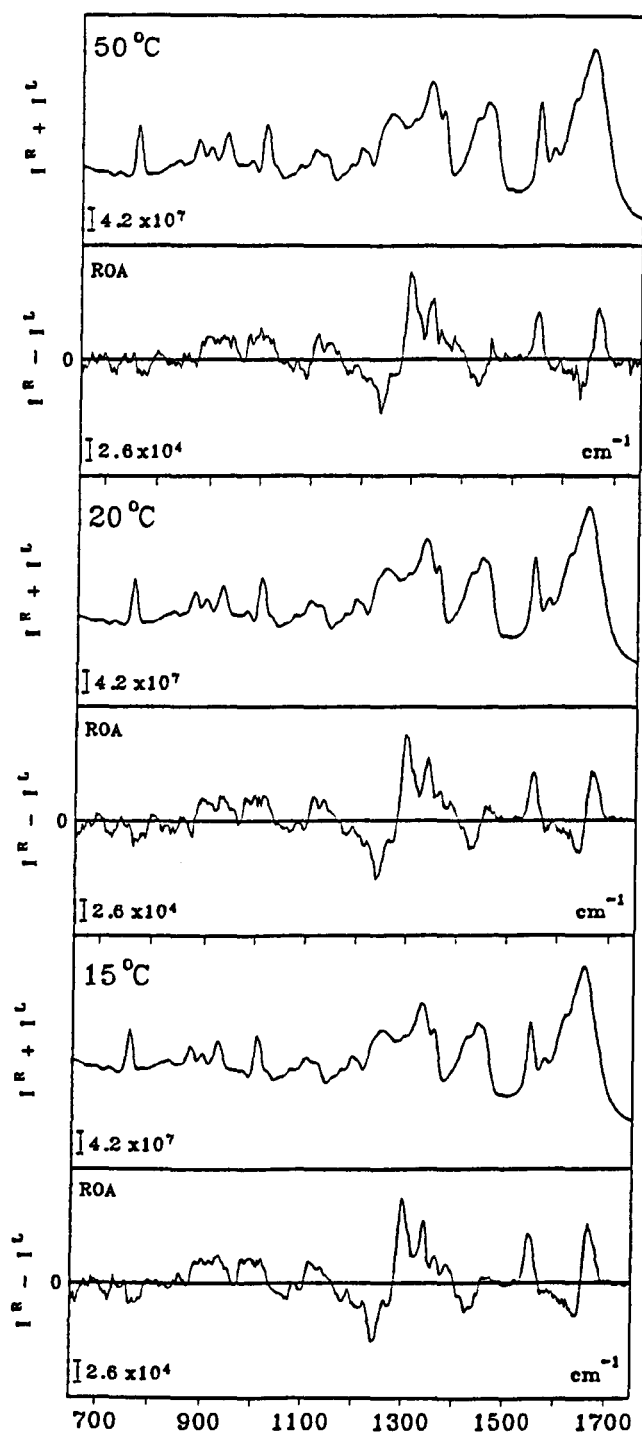


Figure 7.14 Backscattered Raman and ROA of native lysozyme at 50, 20 and 15°C



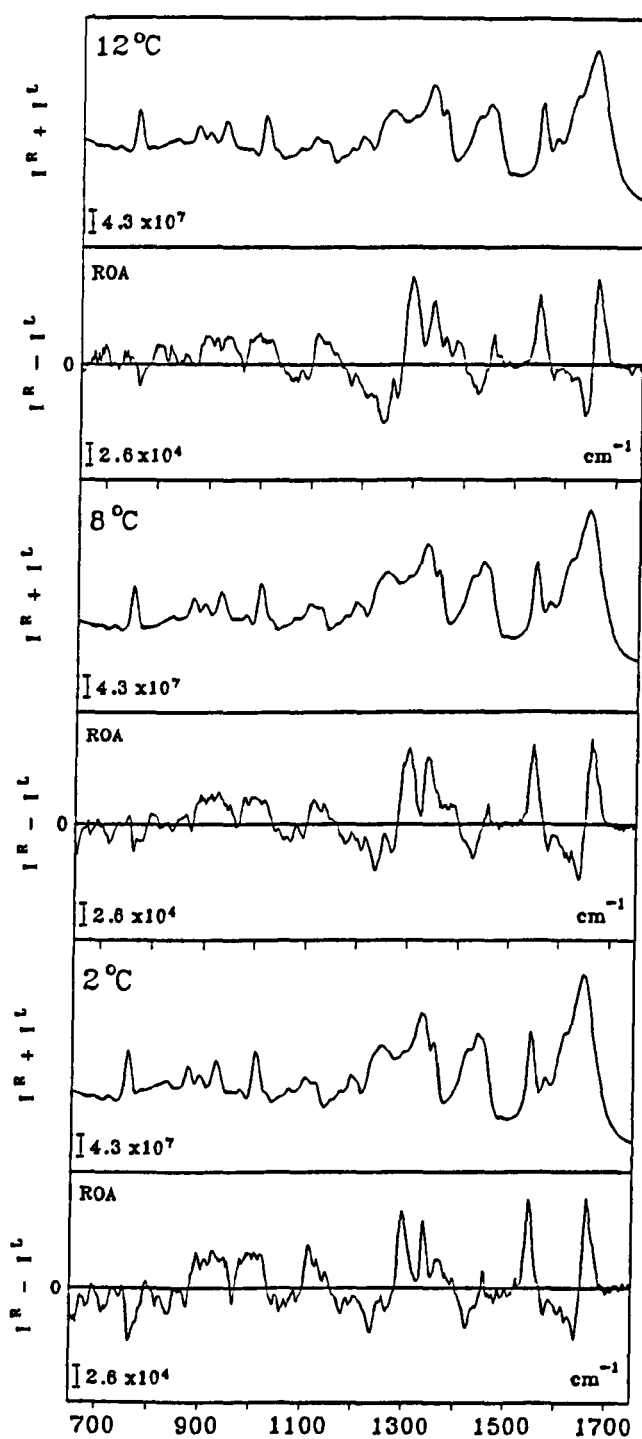


Figure 7.15 Backscattered Raman and ROA of native lysozyme at 12, 8 and 2°C

One of the largest changes that occurs is located in the amide III region. As was previously discussed in reduced lysozyme, the ROA spectrum was dominated by a large negative-positive amide III couplet which was broad and relatively unstructured. This is a consequence of the very fast timescale of Raman spectroscopy i.e. it is much faster than the conformational changes within the protein. Since this broad unstructured couplet is generated by the superposition of 'snapshot' amide III signals (due to the enormous amount of conformers present in the unfolded state) it can be proposed that the couplet which underlies many native protein spectra in this region could be due to parts of the structure with similar conformational heterogeneity and mobility as that of the unfolded protein. At low temperatures this underlying amide III couplet appears to decrease rapidly in intensity; many of the other ROA bands show an accompanying increase in sharpness and intensity. This leads to the suggestion that disordered mobile structure is coalescing into rigid structure. It should also be said at this point that although the ROA spectra reveal structural changes the conventional Raman spectra are virtually the same over all the measured temperatures (if the temperature was raised high enough, which is not possible with the current instrument, the Raman spectra would be observed to change when the protein was denatured by heat).

One of the most interesting bands is the  $1340\text{ cm}^{-1}$  positive band which is thought to be due to  $3_{10}$ -helix loop structure. On lowering the temperature from 20 to  $2^{\circ}\text{C}$  this band increases by  $\sim 40\%$ . In Figure 7.16 there is an illustration of lysozyme in a particular orientation that is helpful in this situation.

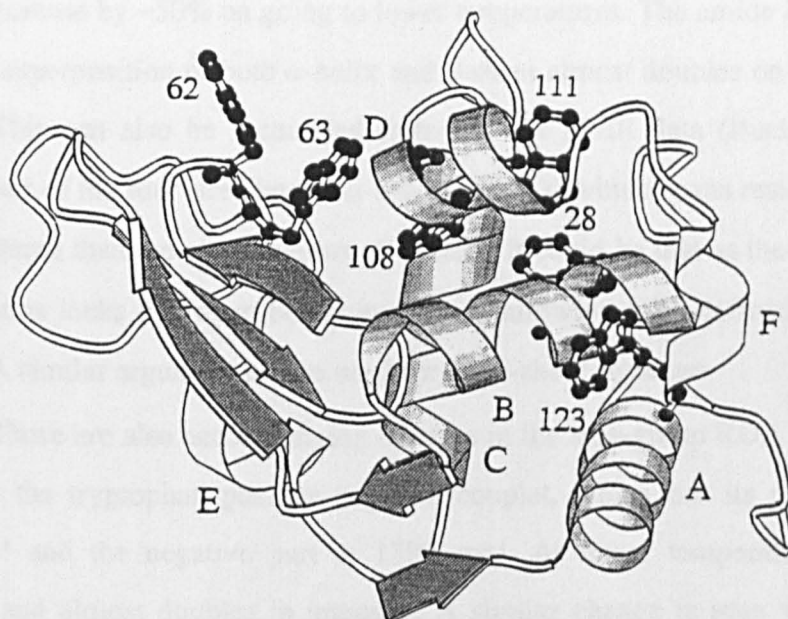


Figure 7.16 Molscript drawing of lysozyme

In the crystal structure of lysozyme (Blake *et al.*, 1965) there are two stretches of  $3_{10}$ -helix, which are labelled as E and F in the Molscript diagram (i.e. residues 79 - 84 and 118 - 124). It is of relevance that in the solution structure determined by 2-dimensional NMR (Smith *et al.*, 1993),  $3_{10}$ -helix was not defined in the 79 - 84 region at the experimental temperature of 35°C (it was assumed by the authors that this was due to a lack of NOE's in this region). This conflict was partly resolved with new  $^{15}\text{N}$  NMR relaxation studies (Buck *et al.*, 1995) which revealed that residues 83 - 88 have low order parameters (especially residue 85) in solution even though the crystal B-factors are of an expected value. This leads to the conclusion that as the temperature is lowered 'mobile' structure is coalescing into nascent  $3_{10}$ -helix. It is more than just a coincidence that the  $1340\text{ cm}^{-1}$  also increases by ~40% when the trimer of N-acetylglucosamine is bound to lysozyme as is shown in Figure 7.13, as binding the saccharide inhibitor induces a much more rigid structure.

It would be disturbing if no other changes were observed in the spectra, but

both the  $\alpha$ -helix ( $\sim 890 - 960 \text{ cm}^{-1}$ ) and  $\beta$ -sheet ( $\sim 1020 - 1060 \text{ cm}^{-1}$ ) positive ROA regions increase by  $\sim 50\%$  on going to lower temperatures. The amide I band which is mainly a superposition of both  $\alpha$ -helix and  $\beta$ -sheet almost doubles on going from 20 to  $2^\circ\text{C}$ . This can also be reconciled from the  $^{15}\text{N}$  NMR data (Buck *et al.*, 1995) because one of the four stretches of  $\alpha$ -helix (helix D) which spans residues 108 - 115 is less ordered than the other regions of  $\alpha$ -helix. It could be that as the temperature is lowered, this locks hydrogen bonds into place, allowing fully-defined  $\alpha$ -helix to be created. A similar argument can be used for the  $\beta$ -sheet structure.

There are also accompanying changes in the side-group ROA band structure involving the tryptophan positive-negative couplet, which has its positive part at  $1555 \text{ cm}^{-1}$  and the negative part at  $1580 \text{ cm}^{-1}$ . At lower temperatures this band sharpens and almost doubles in intensity. A similar change is seen when  $\text{NAG}_3$  is bound to lysozyme, although it is not so large there. There are six tryptophan residues in lysozyme (28, 62, 63, 108, 111 and 123) and the ROA will probe their local geometry relative to the peptide backbone *via* coupling of the vibrational coordinates of the residue with those of the connecting bonds. Due to the increase in sharpness and intensity of the band it suggests that several of the tryptophans are experiencing an increase in rigidity at lower temperatures. It could be relevant to this proposal that tryptophans 62 and 63 have indole hydrogens which are not hydrogen bonded in the crystal. Tryptophan 62 also has a very low solution order parameter from  $^{15}\text{N}$  NMR accompanied by a high B-factor in the crystal (Buck *et al.*, 1995). It is also known that these two tryptophans along with tryptophan 108 are involved in the catalytic function of the protein and from the crystal structure of lysozyme- $\text{NAG}_3$  (Chetham *et al.*, 1992) it was found that tryptophans 62 and 63 moved inwards by  $\sim 0.5$  to  $1.0 \text{ \AA}$ . This will have the effect of closing the active site and allowing hydrogen bonded networks to form, confirmed by much lower B-factors. The tryptophans 62 and 63 are closely linked by the disulfide between cysteines 64 and 80 to the mobile stretch of  $3_{10}$ -helix between residues 79 - 85. Reducing the temperature might therefore have a similar effect of stiffening the structure as was observed in the  $\text{NAG}_3$ .

No discernible change in the wavelength or intensity of the tryptophan fluorescence over the corresponding temperature range was found as is shown in Figure 7.17.

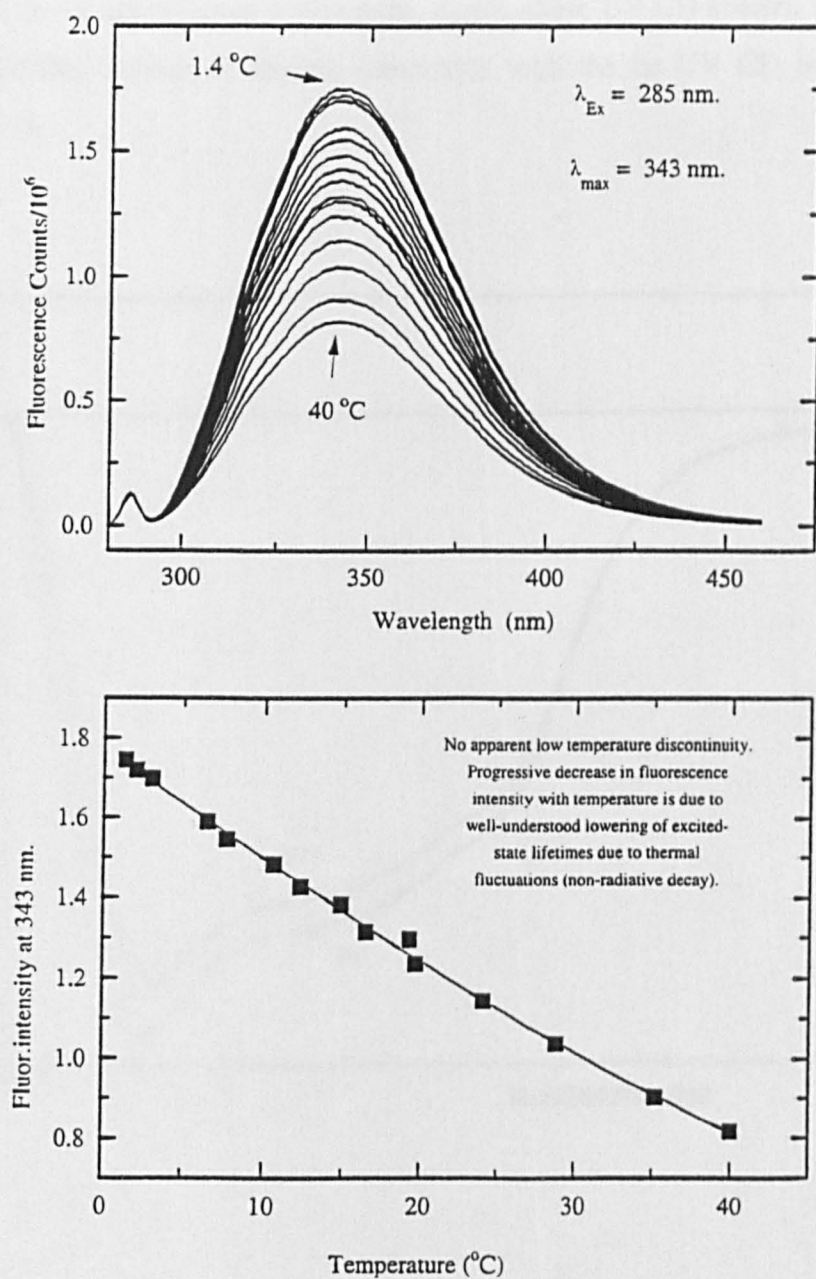


Figure 7.17 Fluorescence of native lysozyme over a range of temperatures

It was also observed that there were no changes in the near-UV CD, indicating that the global environments of the tryptophans have not changed. It was reassuring though to observe a change in the far-UV CD, where an  $\sim 10\%$  increase was observed for the negative  $\alpha$ -helix band at 208 nm; there is also a small shift of this band of  $\sim 1$  nm to lower wavelength. Again, these UV CD spectra were measured at the BBSRC service at Stirling University, with the far-UV CD being shown in Figure 7.18.

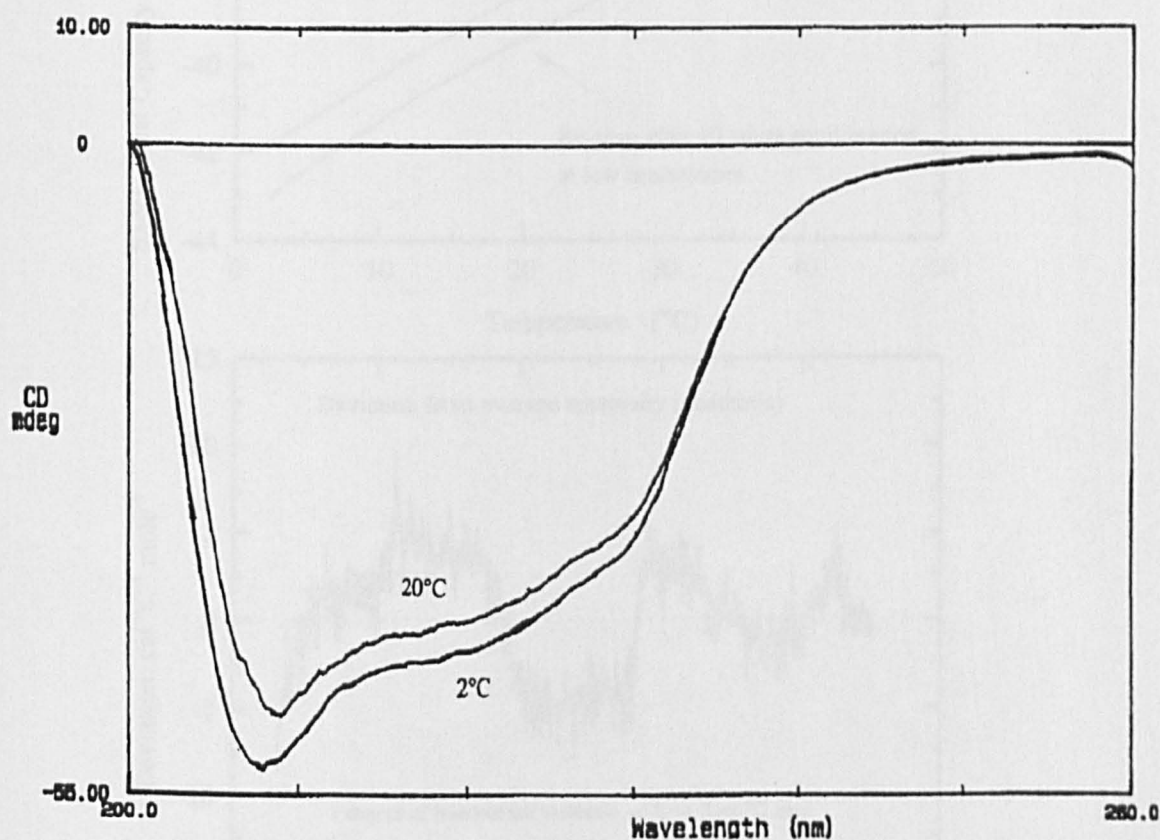


Figure 7.18 UV CD of lysozyme at 20 and 2°C



When DSC measurements on lysozyme were performed down to 2°C, it was worrying that there was no sign of any thermal transition, as shown in Figure 7.19.

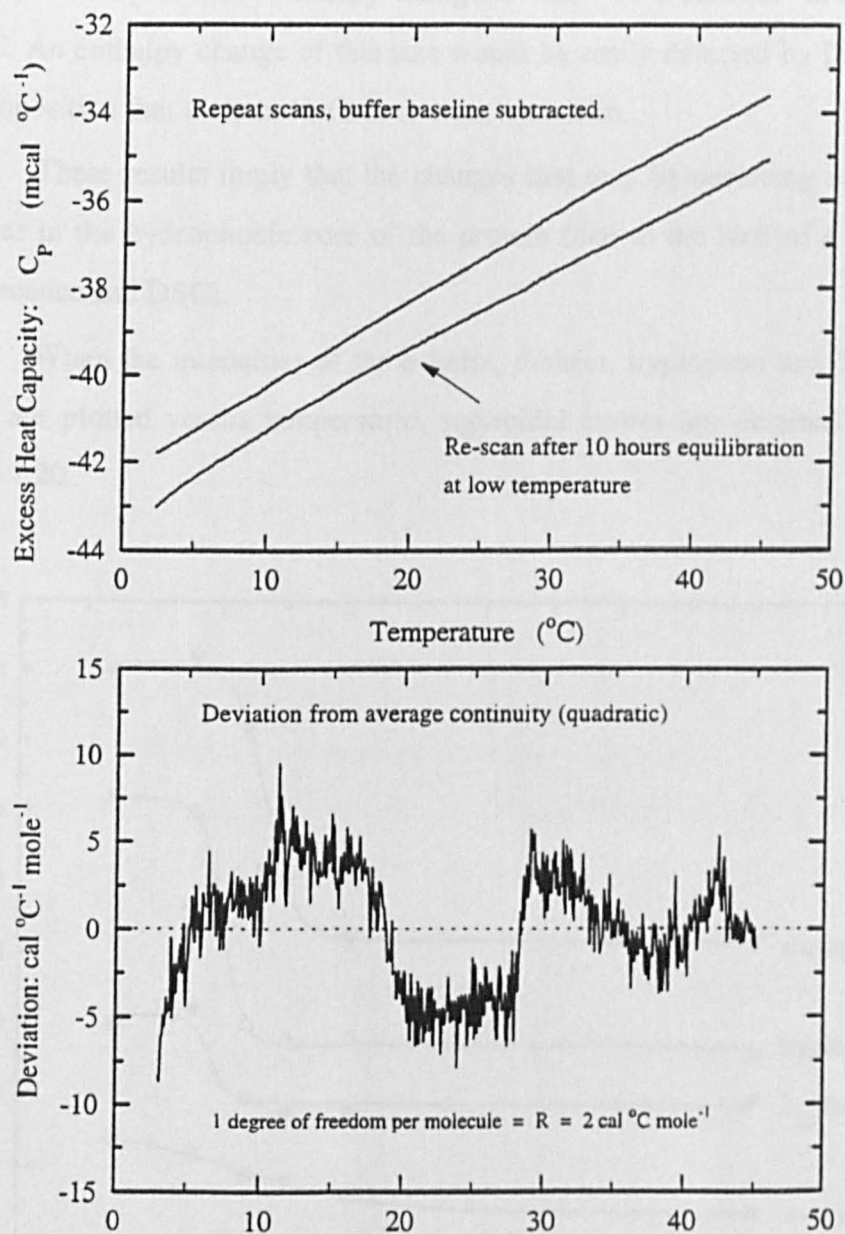


Figure 7.19 DSC of lysozyme over a range of temperatures

If a first-order transition was assumed and the amide I and tryptophan ROA couplets were used in a traditional van't Hoff analysis by plotting  $\ln K$  ( $K$  is the associated equilibrium constant) versus  $1/T$  a straight line plot is obtained. This analysis would predict an enthalpy change of  $\Delta H \approx 90 \text{ k cal mol}^{-1}$  at a transition of  $\sim 12^\circ\text{C}$ . An enthalpy change of this size would be easily detected by DSC, leading to the proposition that it cannot be a first order transition.

These results imply that the changes that may be occurring are restricted to residues in the hydrophobic core of the protein (due to the lack of a change in the fluorescence and DSC).

When the intensities of the  $\alpha$ -helix,  $\beta$ -sheet, tryptophan and  $3_{10}$ -helix ROA bands are plotted versus temperature, sigmoidal curves are obtained as shown in Figure 7.20.

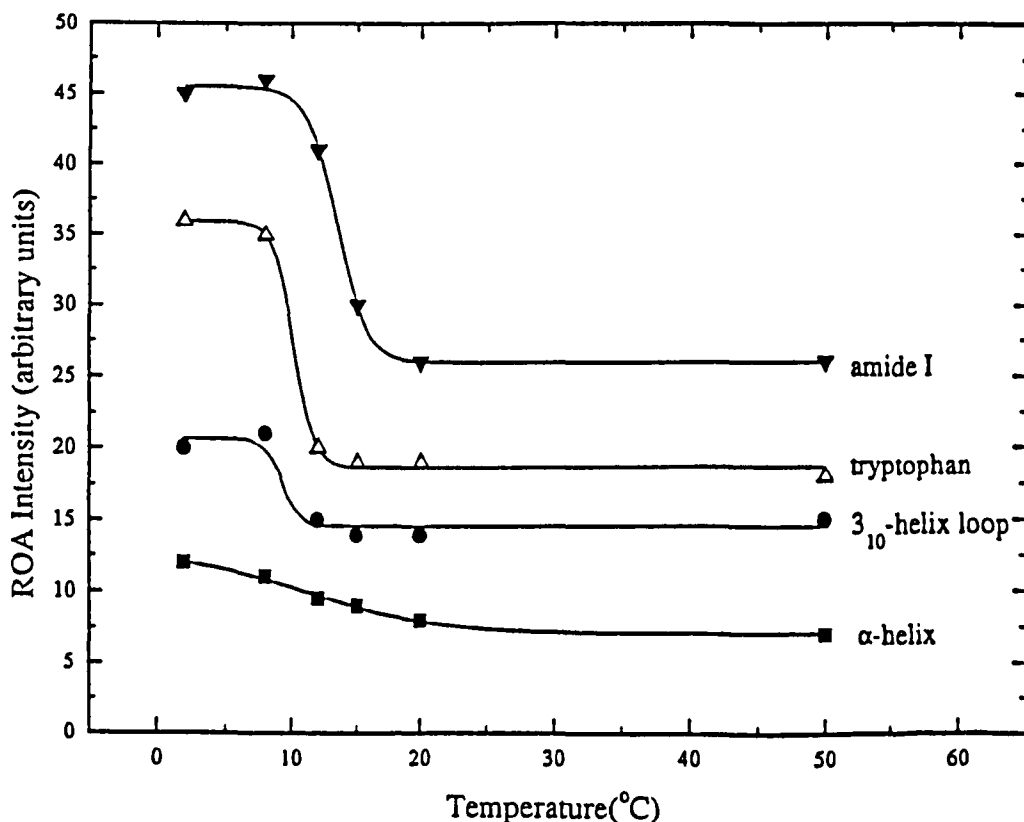


Figure 7.20 Plot of ROA intensity for  $\alpha$ -helix,  $\beta$ -sheet, tryptophan and  $3_{10}$ -helix against temperature



These observations might be related to the glasslike transitions that are known to occur in some proteins. The temperatures of these transitions are however much lower than that described here: from X-ray B-factors, inelastic neutron scattering and Mossbauer spectroscopy these transitions have been found to occur at 180 - 220K (Gregory, 1995). Although the temperatures of the two different types of transitions are quite different, it was observed from a recent X-ray study of hen egg-white lysozyme that there was a much smaller (~50%) number of ordered water molecules in the nearest hydration shell (i.e. 3Å away from the protein) in the structure at 280K, compared to those at 295 or 250K (Kurinov & Harrison, 1995) which is actually close to the transition that has been identified in these ROA experiments. The temperature of glasslike transitions increases as the degree of hydration decreases and at a low hydration level of 0.12h (g water / g protein) the glasslike transition of lysozyme has been observed from positron annihilation lifetime measurements to occur at 298K (Gregory, 1995). A major difference between these glasslike transitions and the ROA transitions, is that the former occur in the solid phase whereas ROA is measured in an aqueous solution.

From the plots of the intensities versus temperature, ROA shows the transition to be cooperative, and since no latent heat is observed (from DSC) it might be a continuous phase transition of the order-disorder type. Residual mobility is therefore freezing into native polypeptide backbone structure. There could be cooperative hydrogen bonded networks that surround the backbone and side-groups and with this concept an analogy can be made with ferroelectric phase transitions (Lines & Glass, 1977). These ferroelectric phase transitions are of the order-disorder type in crystals and involve the quenching of cooperative proton tunneling (Kearley *et al.*, 1994). Evidence for correlated proton tunneling in proteins comes from recent inelastic neutron scattering studies which studied crystalline amides (e.g. N-methylacetamide and polyglycine I) and monitored the vibrations of hydrogen bonds. These experiments showed that the proton is delocalised in two equivalent positions, generating a symmetric double-minimum potential (Kearley *et al.*, 1994). Enzymatic reaction rates have also been shown to be dependent on hydrogen tunneling (Cha *et*

*al.*, 1989). It is known that interactions between protons and water molecules increase the effective proton mass which has the effect of suppressing the tunneling; however, if the correlated proton tunneling in proteins is confined to the hydrophobic core which is similar to the situation in crystals (i.e. no solvent water molecules) there is no problem with the suppression of the tunneling. This could explain why there is no tryptophan fluorescence or near-UV CD changes, although a transition is observed in ROA.

These ROA observations support the current view that "proteins should be treated as a peculiar and interesting class of dielectric glass with regions of higher and lower quasi-degenerate conformational state densities and hence with lower and higher vitrification temperatures" (Gol'danskii *et al.*, 1989). It also raises the question if the physics of spin glasses can be used to understand biological systems (Stein, 1992).

There are however a few problems associated with this experiment which will now be discussed briefly. ROA requires the use of rather high protein concentrations (~70 mg/ml) and therefore there is always the worry that aggregation may be causing the ROA effects. This is especially true when the temperature is varied as here (and also in the molten globule  $\alpha$ -lactalbumin). By monitoring the Rayleigh line and changing the temperature of the lysozyme there was no significant change over the 2 - 50°C temperature range. If significant aggregation was occurring the Rayleigh scatter would increase dramatically.

Another serious problem connected with the temperature dependent lysozyme study is the repeatability of the results. Sometimes the transition is observed and other times when all of the experimental parameters appear to be the same, nothing is observed. This initially was put down to unknown factors until exactly the same situation was found in the far-UV CD experiments where the transition again was not always observed. This raises the possibility that the transition might have some sort of metastable characteristics which is known to occur in some continuous transitions (Sewell, 1986).

## Conclusion

ROA has provided a fresh insight into the dynamic aspects of protein structure. Even with the extensive studies by both NMR and X-ray crystallography on both  $\alpha$ -lactalbumin and lysozyme, ROA has still managed to reveal new aspects in particular about the molten globule states of  $\alpha$ -lactalbumin and as to how the lysozyme structure can be made more rigid by binding an inhibitor or lowering the temperature. ROA has also highlighted the importance of  $\phi$ ,  $\psi$  propensities in unfolded proteins.

## Appendix

### ROA Signature Maps

To summarize and bring together a few of the ideas in the thesis tentative signature maps for both secondary structure and loops and turns will now be proposed.

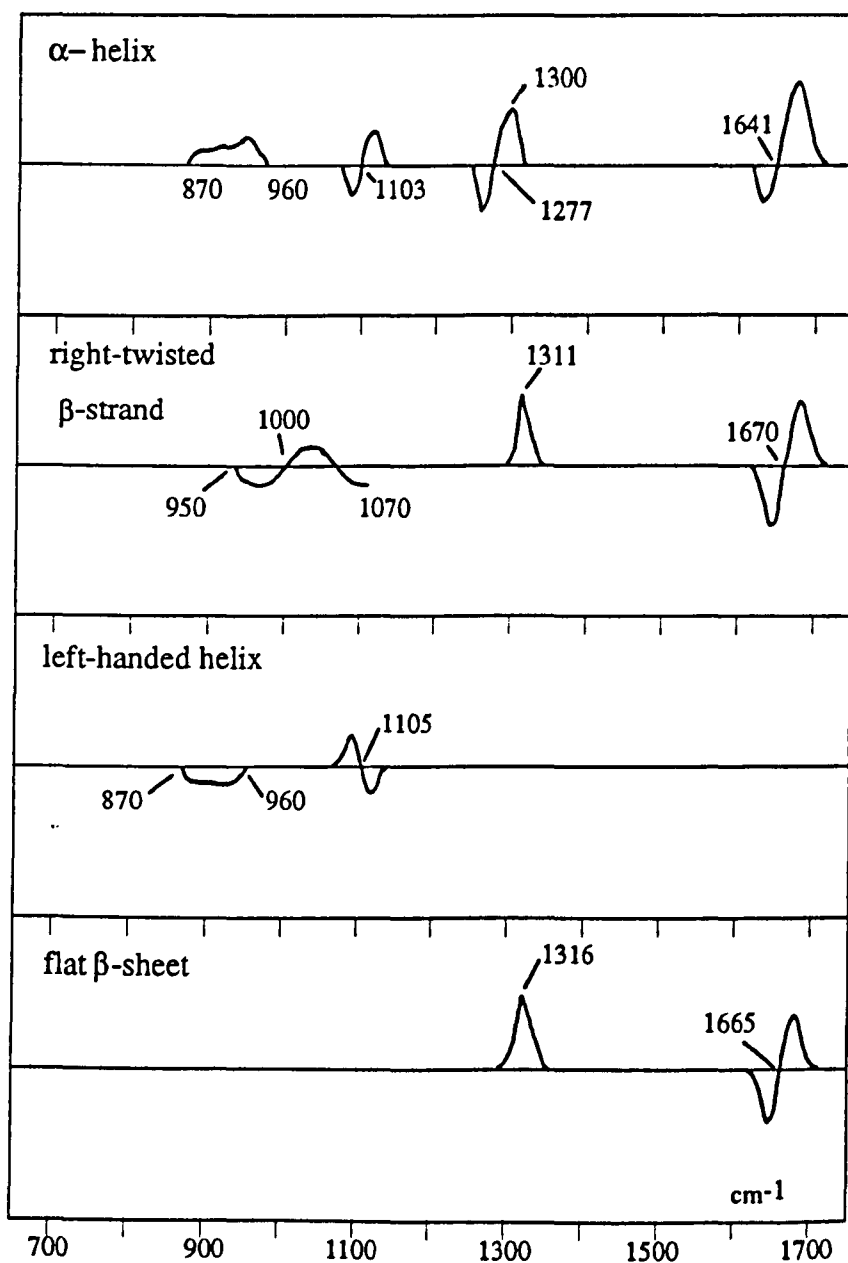


Figure 8.1 ROA signature map for secondary structure

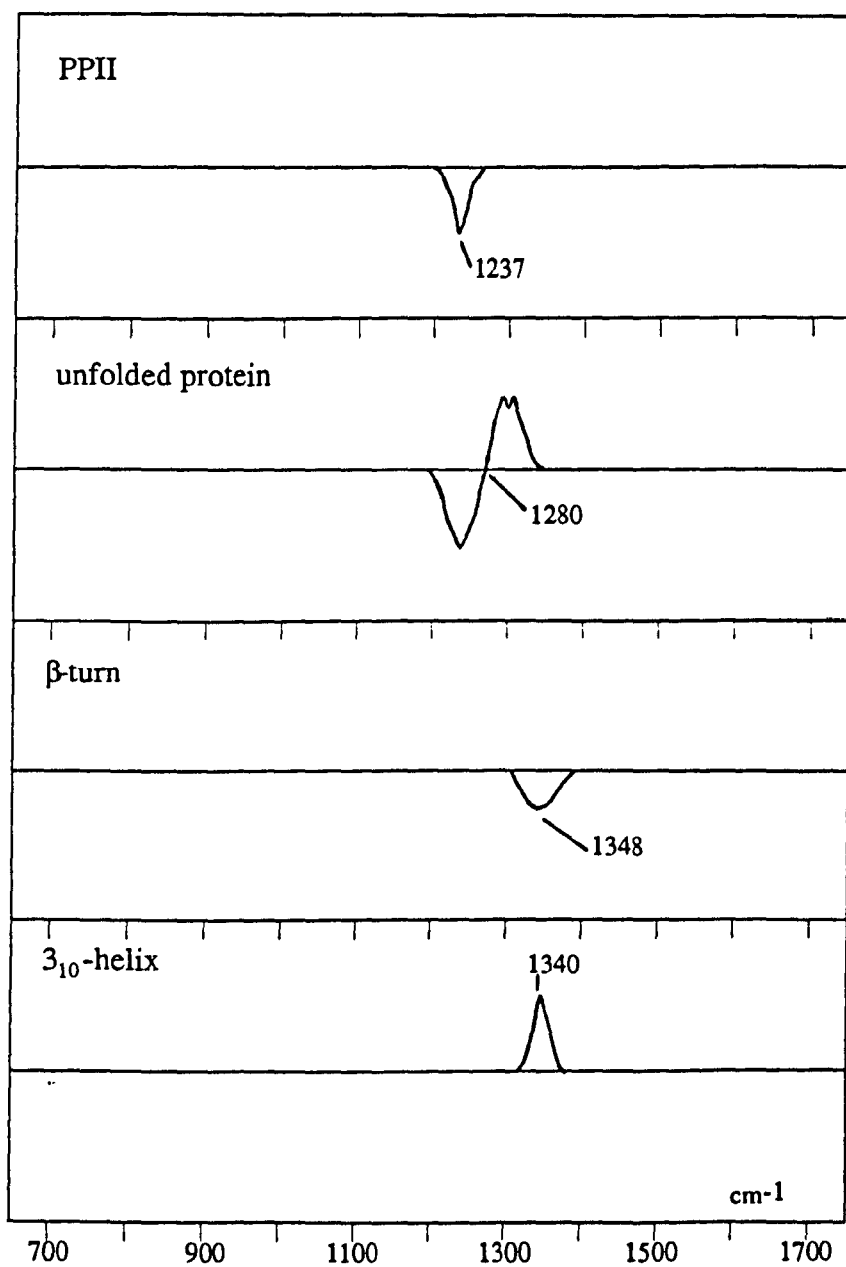


Figure 8.2 ROA signature map of loop and turn structure

## References

- Acharya, K. R., Stuart, D. I., Walker, N. P. C., Lewis, M. & Phillips, D. C. (1989). *J. Mol. Biol.* **208**, 99.
- Acharya, K. R., Ren, J., Stuart, D. I., Phillips, D. C. & Fenna, R. E. (1991). *J. Mol. Biol.* **221**, 571.
- Adzhubei, A. A., Eisenmenger, F., Tumanyan, V. G., Zinke, M., Brodzinski, S. & Esipova, N. G. (1987). *J. Biomol. Strut. Dynam.* **5**, 689.
- Adzhubei, A. A. & Sternberg, M. J. E. (1993). *J. Mol. Biol.* **229**, 472.
- Alexandrescu, A. T., Broadhurst, R. W., Wormald, C., Chyan, C. L., Baum, J. & Dobson, C. M. (1992). *Eur. J. Biochem.* **210**, 699.
- Alexandrescu, A. T., Evans, P. A., Pitkeathly, M., Baum, J. & Dobson, C. M. (1993). *Biochemistry*, **32**, 1707.
- Amos, R. D. (1982). *Chem. Phys. Lett.* **87**, 23.
- Amos, R. D. & Rice, J. E. (1987). 'CADPAC': The Cambridge Analytical Derivatives Package, Cambridge.
- Arago, D. F. J. (1811). *Mém. de l' Inst.* **12**, part 1, 93.
- Atkins, P. W. & Barron, L. D. (1969). *Mol. Phys.* **16**, 453.
- Austin, R. H. & Chen, C. M. (1992). The spin-glass analogy in protein dynamics. In *Spin Glasses in Biology* (Stein, D. L. ed), pp. 179 - 223, World Scientific, Singapore.
- Avey, H. P. M. O., Carlisle, C. H., Evans, S. A., Morris, S. J., Palmer, R. A., Woolhouse, B. A. & Shall, S. (1967). *Nature (London)*, **213**, 557.
- Badger, J. *et al.* (1991). *Acta Cryst.* **B47**, 127.
- Baker, E. N. *et al.* (1988). *Phil. Trans. R. Soc.* **B319**, 389.
- Bandekar, J. (1991). *Biochem. Biophys. Acta*, **123**, 1120.
- Barron, L. D. (1969). Doctoral Thesis, Oxford University, 1969.
- Barron, L. D. & Buckingham, A. D. (1971). *Mol. Phys.* **20**, 1111.

- Barron, L. D., Bogaard, M. P. & Buckingham, A. D. (1973a). *J. Am. Chem. Soc.* **95**, 604.
- Barron, L. D., Bogaard, M. P. & Buckingham, A. D. (1973b). *J. Chem. Soc., Chem. Commun.*, 152.
- Barron, L. D., Bogaard, M. P. & Buckingham, A. D. (1973c). *Nature (London)*, **241**, 113.
- Barron, L. D. & Buckingham, A. D. (1974). *J. Am. Chem. Soc.* **96**, 4769.
- Barron, L. D. & Buckingham, A. D. (1975). *Annu. Rev. Phys. Chem.* **26**, 381.
- Barron, L. D. & Buckingham, A. D. (1979). *J. Am. Chem. Soc.* **101**, 1979.
- Barron, L. D. (1982). *Molecular Light Scattering and Optical Activity*, Cambridge University Press, Cambridge.
- Barron, L. D. & Clark, B. P. (1982). *Mol. Phys.* **46**, 839.
- Barron, L. D., Hecht, L., Hug, W. & MacIntosh, M. J. (1989). *J. Am. Chem. Soc.* **111**, 8731.
- Barron, L. D., Gargaro, A. R. & Wen, Z. Q. (1990). *J. Chem. Soc.* 1034.
- Barron, L. D., Gargaro, A. R., Hecht, L. & Polavarapu, P. L. (1991). *Spectrochim. Acta.* **47A**, 1001.
- Barron, L. D., Gargaro, A. R., Hecht, L. & Polavarapu, P. L. (1992). *Spectrochim. Acta.* **48A**, 261.
- Barron, L. D., Hecht, L., Gargaro, A. R. & Hug, W. (1992). *J. Raman Spectrosc.* **21**, 375.
- Barron, L. D., Cooper, A., Ford, S. J., Hecht, L. & Wen, Z. Q. (1992a). *Faraday Discuss.* **93**, 259.
- Barron, L. D., Hecht, L., (1993). In *Advances in Spectroscopy*, (Clark, R. J. H. & Hester, R. E. eds), vol. 21, pp. 235 - 266, Wiley, Chichester.
- Barron, L. D., Ford, S. J., Bell, A. F., Wilson, G., Hecht, L. & Cooper, A. (1995). *Faraday Discuss.* **99**, 217.
- Barron, L. D., Hecht, L., Bell, A. F. & Wilson, G. (1996). *Applied Spectroscopy*, in press.

- Bartunik, H. D., Summers, L. J. & Bartsch, H. H. (1989). *J. Mol. Biol.* **210**, 813.
- Batley, D. E., Slater, J. B., Wludyka, R., Owen, H., Pallister, D. M. & Morris, M. D. (1993). *Appl. Spectrosc.* **47**, 1913.
- Becker, J. W., Reeke, G. N., Wang, J. L., Cunningham, B. A. & Edelman, G. M. (1975). *J. Biol. Chem.* **250**, 1513.
- Bell, A. F., Hecht, L. & Barron, L. D. (1993). *J. Raman Spectrosc.* **24**, 633.
- Bell, A. F. (1994). Doctoral Thesis, Glasgow University, 1994.
- Bell, A. F., Hecht, L. & Barron, L. D. (1994). *J. Am. Chem. Soc.* **116**, 5155.
- Bentley, G., Dodson, E., Dodson, G., Hodgkin, D. & Mercola, D. (1976). *Nature*, **261**, 166.
- Bernard, M., Canioni, P., Cozzzone, P., Berthou, J. & Jolles, P. (1990). *Int. J. Peptide Protein Chem.* **36**, 46.
- Berliner, L. J., Koga, K., Nishikawa, H. & Scheffler, J. E. (1987). *Biochemistry*, **26**, 5769.
- Bertouzza, A., Bonora, S., Fini, G. & Morelli, M. A. (1992). *J. Appl. Spectrosc.* **37**, 58.
- Bickel, W. S. & Bailey, W. M. (1985). *Am. J. Phys.* **53**, 468.
- Bilhorn, R. B., Sweedler, J. V., Epperson, P. M. & Denton, M. B. (1987a). *Appl. Spectrosc.* **41**, 1114.
- Bilhorn, R. B., Epperson, P. M., Sweedler, J. V. & Denton, M. B. (1987b). *Appl. Spectrosc.* **41**, 1125.
- Binney, J. J., Dowrick, N. J., Fisher, A. J. & Newman, M. E. J. (1992). *The Theory of Critical Phenomena*, Clarendon Press, Oxford.
- Biot, J. B. (1812). *Mém, de l' Inst.* **13**, part 1, 218.
- Biot, J. B. (1815). *Bull. Soc. Philomath.* 190.
- Birke, S. S., Agbaje, I. & Diem, M. (1991). *Biochemistry*, **31**, 450.
- Blake, C. C. F., Koenig, D. F., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, R. (1965). *Nature (London)*, **206**, 757.



- Blake, C. C. F., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. (1967). *Proc. R. Soc. London, Ser. B.* **167**, 365.
- Blake, C. C. F., Cassels, R., Dobson, C. M., Poulsen, F. M., Williams, R. J. P. & Wilson, K. S. (1981). *J. Mol. Biol.* **147**, 73.
- Blout, E. R. & Fasman, G. D. (1958). In *Recent Advances in Gelatin and Glue Research* (Stainsby, G. ed), pp. 122 - 130, Pergamon, New York.
- Blundell, T. E. & Hubbel, R. E. (1980). *Nature (London)* **287**, 781.
- Bode, W. & Schwager, P. (1975). *J. Mol. Biol.* **98**, 693.
- Bode, W. & Huber, R. (1978). *Acc. Chem. Res.* **11**, 114.
- Borkakoti, N., Moss, D. S. & Palmer, R. A. (1982). *Acta. Cryst. Sect. B*, **38**, 2210.
- Borman, S. (1995). *Chemical and Engineering News*, November 6, 20.
- Bose, P. K., Polavarapu, P. L., Barron, L. D. & Hecht, L. (1990). *J. Phys. Chem.* **94**, 1734.
- Bosnich, B., Moskovits, M. & Ozin, G. A. (1972). *J. Am. Chem. Soc.* **94**, 4750.
- Bradbury, J. H. & Wilairiat, P. (1967). *Biochem. Biophys. Res. Commun.* **29**, 84.
- Bradbury, J. H. & Brown, L. R. (1977). *Biochemistry*, **16**, 573.
- Bradbury, J. H., Ramesh, R. & Dodson, G. (1981). *J. Mol. Biol.* **150**, 609.
- Branden, C. & Tooze, J. (1991). *Introduction to Protein Structure*, Garland Publishing, New York.
- Brooks, C. L., Karplus, M. & Pettitt, B. M. (1988). *Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics*. (Volume 71 of *Adv. Chem. Phys.*).
- Brown, J. R. (1975). *Fed. Proc., Fed. Am. Soc. Exp. Biol.* **34**, 591.
- Brown, J. E. & Klee, W. A. (1971). *Biochemistry*, **10**, 470.
- Buck, M., Boyd, J., Redfield, C., Mac Kenzie, D. A., Jeenes, D. J., Archer, D. B. & Dobson, C. M. (1995). *Biochemistry*, **34**, 4041.
- Buckingham, A. D. & Longuet-Higgins, H. C. (1968). *Mol. Phys.* **14**, 63.

- Buckingham, A. D. & Dunn, M. B. (1971). *J. Chem. Soc. A*, 1988.
- Campbell, R. L. & Petsko, G. A. (1987). *Acta. Cryst. Sect. B*, **38**, 2210.
- Carey, P. R. (1982). *Biochemical Applications of Raman and Resonance Raman Spectroscopies*, Academic Press, New York.
- Carter, D. C. *et al.* (1989). *Science*, **244**, 1195.
- Carter, D. C. & He, X. M. (1990). *Science*, **249**, 302.
- Cahill, R. Jr. (1971). *Diabetes*, **20**, 785.
- Cha, Y., Murray, C. J. & Klinman, J. P. (1989). *Science*, **243**, 1325.
- Chan, H. S. & Dill, K. A. (1993). *Physics Today*, February, 24.
- Chan, H. S., Bronberg, S. & Dill, K. A. (1995). *Phil. Trans. Roy. Soc. Lond.* **B348**, 61.
- Che, D., Hecht, L. & Nafie, L. A. (1991). *Chem. Phys. Lett.* **180**, 182.
- Che, D. & Nafie, L. A. (1992). *Chem. Phys. Lett.* **189**, 35.
- Cheetham, J. C., Artymiuk, P. J. & Phillips, D. C. (1992). *J. Mol. Biol.* **224**, 613.
- Chen, M. C. & Lord, R. C. (1976). *J. Am. Chem. Soc.* **98**, 990.
- Christensen, H. & Pain, R. H. (1994). *Mechanisms of Protein Folding*. (Pain, R. H. ed), pp. 55 - 79, IRL Press, Oxford.
- Chothia, C., Lesk, A. M., Dodson, G. G. & Hodgkin, D. C. (1983). *Nature*, **302**, 500.
- Chyan, C. - L., Wormald, C., Dobson, C. M., Evans, P. A. & Baum, J. (1993). *Biochemistry*, **32**, 5681.
- Clore, G. M., Wingfield, P. T. & Gronenborn, A. (1991). *Biochemistry*, **30**, 2315.
- Colthup, N. B., Daley, L. H. & Wiberley, S. E. (1975). *Introduction to Infrared and Raman Spectroscopy*, Academic Press, New York, 1975.
- Cooper, J. B., Khan, G., Taylor, G., Tickle, I. J. & Blundell, T. L. (1990). *J. Mol. Biol.* **214**, 199.
- Cotton, A. (1895). *Compt. Rend.* **120**, 989.

- Cowan, P. M. & Mc Gavin, S. (1955). *Nature*, **176**, 501.
- Creighton, T. E. (1988). *Biophys. Chem.* **31**, 155.
- Creighton, T. E. (1989). In *Protein Structure: a Practical Approach*, (T. E. Creighton ed), pp. 155, IRL Press, Oxford.
- Creighton, T. E. (1992). *Protein Folding*, W. H. Freeman & Co., New York.
- Creighton, T. E. (1994). *Proteins*, W. H. Freeman & Co., New York.
- De Meyts, P., Van Obberghen, E., Roth, J., Brandenburg, D. & Wollmer, A. (1978). *Nature (London)*, **273**, 504.
- Dearborn, D. G., Wetlaufer, D. B. (1970). *Biochem. Biophys. Res. Comm.* **39**, 314.
- Dearwenda, U. *et al.* (1989). *Nature*, **338**, 594.
- Diem, M., Fry, J. L. & Burrow, D. F. (1973). *J. Am. Chem. Soc.* **95**, 253.
- Diem, M., Lee, O. & Roberts, G. M. (1992). *J. Phys. Chem.* **96**, 548.
- Diem, M. (1993). *Modern Vibrational Spectroscopy*, Wiley, New York.
- Dierker, S. B., Murray, C. A., Legrange, J. D. & Schlotter, N. E. (1987). *Chem. Phys. Lett.* **137**, 453.
- Dill, K. A. & Shortle, D. (1991). *Annu. Rev. Biochem.* **60**, 795.
- Dill, K. A., Bromberg, S., Yue, K., Fiebig, K. M., Yee, D. P., Thomas, P. D. & Chan, H. S. (1995). *Protein Sci.* **4**, 561.
- Dobson, C. M. (1994). *Curr. Biol.* **4**, 636.
- Dodson, E. J., Dodson, G. G. & Hodgkin, D. C. (1979). *Can. J. Biochem.* **57**, 469.
- Dolgikh, D. A., Gilmanishin, R. I., Brazhnikov, E. V., Bychkova, V. E., Semisotnov, G. V., Venyaminov, S. Y. & Ptitsyn, O. B. (1981). *FEBS Letters*, **136**, 311.
- Dollish, F. R., Fateley, W. G. & Bentley, F. F. (1974). *Characteristic Raman Frequencies of Organic Compounds*, Wiley, New York.
- Doty, P., Wade, A., Yang, J. J. & Blout, E. R. (1957). *J. Poly. Sci.* **23**, 851.
- Drake, A. F., Siligardi, G. & Gibbons, W. A. (1988). *Biophys. Chem.* **31**, 143.

- Dukor, R. K. & Keiderling, T. A. (1991). *Biopolymers*, **31**, 1747.
- Edmundson, A. B., Ely, K. R., Sly, D. A., Westholm, F. A., Powers, D. A. & Liener, I. E. (1971). *Biochemistry*, **10**, 3554.
- Elber, R. & Karplus, M. (1987). *Science*, **235**, 318.
- Epperson, P. M. & Denton, M. B. (1990). *Anal. Chem.* **61**, 1513.
- Escribano, J. R. (1985). *Chem. Phys. Lett.* **121**, 191.
- Escribano, J. R. & Barron, L. D. (1988). *Mol. Phys.* **65**, 327.
- Evans, P. A., Topping, K. D., Woolfson, D. N. & Dobson, C. M. (1991). *Proteins: Structure, Function and Genetics*, **9**, 248.
- Ewbank, J. J. & Creighton, T. E. (1993). *Biochemistry*, **32**, 3694.
- Ewbank, J. J., Creighton, T. E., Hayer - Hartl, M. K. & Hartl, F. U. (1995). *Nature Struct. Biol.* **2**, 10.
- Fairbrother, W. J., Gippert, G. P., Reizer, J., Saier, M. H. & Wright, P. E. (1992). *FEBS Lett.* **296**, 148.
- Fasman, G. D. (1967) In *Poly- $\alpha$ -Amino Acids* (Fasman, G. D. ed), pp. 499, Dekker, New York.
- Fasman, G. D., Itoh, K., Liu, C. S. & Lord, R. C. (1978). *Biopolymers*, **17**, 1729.
- Fehlhammer, H., Bode, W. & Huber, R. (1977). *J. Mol. Biol.* **111**, 415.
- Figge, J., Rossing, T. H. & Fencl, V. (1991). *J. Lab. Clin. Med.* **117**, 453.
- Fink, A. L. (1995). *Ann. Rev. Biophys. Biomol. Struct.* **24**, 495.
- Ford, S. J., Wen, Z. Q., Hecht, L. & Barron, L. D. (1994). *Biopolymers*, **34**, 303.
- Ford, J. S. (1995a). Doctoral Thesis, Glasgow University, 1995.
- Ford, S. J., Cooper, A., Hecht, L., Wilson, G. & Barron, L. D. (1995b). *J. Chem. Soc. Faraday Trans.* **91**, 2087.
- Fox, J. A., Tu, A. T., Hruby, V. J. & Mosberg, H. I. (1981). *Arch. Biochem. Biophys.* **211**, 628.

- Freedman, T. B. & Nafie, L. A. (1994). In *Modern Nonlinear Optics* (Evans, H. & Kielich, S. eds), pp. 207, Wiley, New York.
- Gargaro, A. R., Barron, L. D. & Hecht, L. (1993). *J. Raman Spectrosc.* **24**, 91.
- Gill, S. C. & von Hippel, P. H. (1989). *Anal. Biochem.* **182**, 319.
- Gol'danskii, V. I., Trakhtenberg, L. I. & Fleurov, V. N. (1989). *Tunneling Phenomena in Chemical Physics*, Gordon and Breach, New York.
- Greenfield, N. & Fasman, G. A. (1969). *Biochemistry*, **8**, 4108.
- Gregory, R. B. (1995). In *Protein-Solvent Interactions* (Gregory, R. B. ed), pp. 191, Dekker, New York.
- Griko, Y. V., Freire, E. & Privalov, P. L. (1994). *Biochemistry*, **33**, 1889.
- Grzesiek, S., Döbel, H., Genz, R., Garotta, G., Labhardt, A. H. & Bax, A. (1992). *Biochemistry*, **31**, 8180.
- Gupta, V. P. & Keiderling, T. A. (1992). *Biopolymers*, **32**, 239.
- Günther, H. (1995). *NMR Spectroscopy*, Wiley, Chichester.
- Haas, E., Mc Wherter, C. A. & Sheraga, H. A. (1988). *Biopolymers*, **27**, 1.
- Harada, I. & Takeuchi, H. (1986). In *Advances in Spectroscopy*, (Clark, R. J. H. & Hester, R. E. eds), vol. 13, pp. 113 - 175, Wiley, Chichester.
- Hayer-Hartl, M. K., Ewbank, J. J., Creighton, T. E. & Hartl, F. U. (1994). *EMBO J.* **13**, 3192.
- He, X. M. & Cater, D. C. (1992). *Nature (London)*, **358**, 209.
- Hecht, L. & Jordanov, B. (1987). *Optik*, **75**, 167.
- Hecht, L., Barron, L. D. & Hug, W. (1989). *Chem. Phys. Lett.* **158**, 341.
- Hecht, L. & Barron, L. D. (1990). *Appl. Spectrosc.* **44**, 483.
- Hecht, L., Che, D. & Nafie, L. A. (1991). *Appl. Spectrosc.* **45**, 18.
- Hecht, L. & Nafie, L. A. (1991). *Mol. Phys.* **72**, 441.

- Hecht, L., Barron, L. D., Gargaro, A. R., Wen, Z. Q. & Hug, W. (1992). *J. Raman Spectrosc.* **23**, 401.
- Hecht, L. & Barron, L. D. (1994). *Faraday Discuss.* **99**, 35.
- Higashijima, T., Tasumi, M., Miyazawa, T. & Miyoshi, M. (1978). *Eur. J. Biochem.* **89**, 543.
- Hiltner, W. A., Hopfinger, A. J. & Walton, A. G. (1972). *J. Am Chem. Soc.* **94**, 4324.
- Holowachuk, E. W. (1991). *Gene Bank Database*, Accesion No. M73993.
- Holzwarth, G., Hsu, E. C., Mosher, H. S., Faulkner, T. R. & Moscowitz, A. (1974). *J. Am. Chem. Soc.* **96**, 251.
- Hseu, T. H. & Chang, H. (1980). *Biochim. Biophys. Acta*, **624**, 340.
- Hu, J., Sheng, R. S., Xu, Z. S. & Zeng, Y. (1995). *Spectrochim. Acta Ser. A*, **51**, 1087.
- Iben, I. E. T. *et al.* (1989). *Phys. Rev. Lett.* **62**, 1916.
- Itoh, K., Foxman, B. & Fasman, G. D. (1976). *Biopolymers*, **15**, 419.
- Jacobsen, C. (1972). *Eur. J. Biochem.* **27**, 513.
- Jackson, M., Haris, P. I. & Chapman, D. (1989). *Biochem. et Biophys. Acta*, **998**, 75.
- Jeffreys, H. (1931). In *Cartesian Tensors*, Cambridge University Press, Cambridge.
- Jenness, D. D., Sprecher, C. & Johnson, W. C. Jr. (1976). *Biopolymers*, **15**, 513.
- Kabsch, W. & Sander, M. (1983). *Bioploymers*, **22**, 2557.
- Karpenko, V., Sinkorova, L. & Kodicek, M. (1992). *Coll. Czech. Chem. Commun.* **57**, 641.
- Kartha, G., Bello, J. & Harker, D. (1967). *Nature (London)*, **213**, 862.
- Kearley, G. J., Fillaux, F., Baron, M.-H., Bennington, S. & Tomkinson, J. (1994). *Science*, **264**, 1285.
- Keiderling, T. A. & Pancoska, P. (1993). In *Advances in Spectroscopy* (Clark, R. J. H. & Hester, R. E. eds), vol. 21, pp. 267 - 315, Wiley, Chichester.

- Keiderling, T. A., Wang, B., Urbanova, M., Pancoska, P. & Dukor, R. K. (1994). *Faraday Discuss.* **99**, 263.
- Kiefhaber, T., Labhardt, A. M. & Baldwin, R. L. (1995). *Nature*, **375**, 513.
- Kim, P. S. & Baldwin, R. L. (1990). *Annu. Rev. Biochem.* **59**, 631.
- Kirkwood, J. G. (1937). *J. Chem. Phys.* **5**, 479.
- Klinger, D. S., Lewis, J. S. & Randall, C. E. (1990). *Polarized Light in Optics and Spectroscopy*, Academic Press, London.
- Knobloch, H., Duschl, C. & Knoll, W. (1989). *J. Chem. Phys.* **91**, 3810.
- Kossiakoff, A. A., Chambers, J. L., Kay, L. M. & Stroud, R. M. (1977). *Biochemistry*, **16**, 654.
- Kowalsky, A. (1962). *J. Biol. Chem.* **237**, 1807.
- Kraulis, P. J. (1991). *J. Appl. Crystallogr.* **24**, 946.
- Kreevoy, M. M. & Mead, C. A. (1965). *Faraday Discuss.* **39**, 166.
- Krimm, S. & Mark, J. E. (1968). *Proc. Natl. Acad. Sci. USA*, **60**, 1122.
- Krimm, S., Mark, J. E. & Tiffany, M. L. (1969). *Biopolymers*, **8**, 695.
- Krimm, S. & Bandekar, J. (1986). *Advan. Protein Chem.* **38**, 181.
- Krimm, S. & Reisdorf, W. C. Jr. (1994). *Faraday Discuss.* **99**, 18.
- Kronman, M. J., Sinha, S. K. & Brew, K. (1981). *J. Biol. Chem.* **256**, 8582.
- Kumagai, I., Maenaka, K., Sunada, F., Takeda, S. & Miura, K. (1993). *Eur. J. Biochem.* **212**, 151.
- Kurinov, I. V. & Harrison, R. W. (1995). *Acta. Cryst.* **D51**, 98.
- Kuwajima, K. (1989). *Proteins: Struct. Funct. Genet.* **6**, 87.
- Lacy, W. B., Rowlen, K. L. & Harris, J. M. (1991). *Appl. Spectrosc.* **45**, 1598.
- Lala, A. K. & Kaul, P. (1992). *J. Biol. Chem.* **267**, 19914.

- Landau, L. D. & Lifshitz, L. P. (1980). *Statistical Physics, Part 1* (third edition revised and enlarged by Lifshitz, E. M. & Pitaevskii, L. P.), Pergamon Press, Oxford.
- Lee, D. C., Harris, P. I., Chapman, D. & Mitchell, R. C. (1990). *Biochemistry*, **29**, 9185.
- Lines, M. E. & Glass, A. M. (1977). *Principles and Applications of Ferroelectrics and Related Materials*, Clarendon Press, Oxford.
- Liu, D. J. & Day, L. A. (1994). *Science*, **265**, 671.
- Long, D. A. (1977). *Raman Spectroscopy*, Mc Graw-Hill, New York.
- Lord, R. C. & Yu, N. -T. (1970). *J. Mol. Biol.* **50**, 509.
- Lowe, M. A., Stephens, P. J. & Segal, G. A. (1986). *Chem. Phys. Lett.* **123**, 108.
- Lumb, K. J., Cheetham, J. C. & Dobson, C. M. (1994). *J. Mol. Biol.* **235**, 1072.
- MacCallum, P. H., Poet, R. & Milner-White, E. J. (1995). *J. Mol. Biol.* **248**, 374.
- MacDermott, A. J. & Tranter, G. E. (1989). *Chem. Phys. Lett.* **163**, 1.
- Mantsch, H. H., Casal, H. L. & Jones, R. N. (1986). In *Advances in Spectroscopy*, (Clark, R. J. H. & Hester, R. E. eds), vol. 13, pp. 1 - 42, Wiley, Chichester.
- Marshall, W. E. & Porath, J. (1965). *J. Biol. Chem.* **240**, 209.
- Mc Kenzie, H. A. & White, F. H. (1991). *Advan. Protein Chem.* **41**, 173.
- Mc Reynolds, L., O'Malley, B. W., Nisbet, A. D., Fothergill, J. E., Girol, D., Fields, S., Robertson, L. A. & Brownlee, G. G. (1978). *Nature (London)*, **273**, 723.
- Marquart, M., Walter, J., Deisenhofer, J., Bode, W. & Huber, R. (1983). *Acta Cryst. Sect. B*, **39**, 480.
- Miller, M., Weinstein, J. N. & Wlodawer, A. (1983). *J. Biol. Chem.* **258**, 5864.
- Millhauser, G. L. (1995). *Biochemistry*, **34**, 3873.
- Minor, D. L. & Kim, P. S. (1996). *Nature*, **380**, 730.
- Mirmira, R. & Tager, H. S. (1989). *J. Biol. Chem.* **264**, 6349.
- Mochizuki, K. (1984). *Appl. Opt.* **23**, 3284.



- Murakami, K., Andree, P. J. & Berliner, L. J. (1982). *Biochemistry*, **21**, 5488.
- Murthy, M. R. N., Reid, T. J., Sicignano, A., Tanaka, N. & Rossmann, M. G. (1981). *J. Mol. Biol.* **152**, 465.
- Nafie, L. A., Cheng, J. C. & Stephens, P. J. (1975). *J. Am. Chem. Soc.* **97**, 3842.
- Nafie, L. A., Keiderling, T. A. & Stephens, P. J. (1976). *J. Am. Chem. Soc.* **98**, 2715.
- Nafie, L. A., Diem, M. & Vidrine, D. W. (1979). *J. Am. Chem. Soc.* **101**, 496.
- Nafie, L. A. & Freedman, T. B. (1989). *Chem. Phys. Lett.* **154**, 260.
- Nafie, L. A., Che, D., Yu, G. -S. & Freedman, T. B. (1991). In *Biomolecular Spectroscopy II* (Birge, R. R. & Nafie, L. A. eds), pp. 37 - 40, SPIE, Bellingham, WA.
- Nafie, L. A., Yu, G. -S., Qu, X. & Freedman, T. B. (1994). *Faraday Discuss.* **99**, 13.
- Naik, V. M. & Krimm, S. (1984). *Int. J. Peptide Protein Res.* **23**, 1.
- Newcomer, M. E. *et al.* (1984). *EMBO J.* **3**, 1451.
- Okazaki, A., Ikura, T., Nikaido, K. & Kuwajima, K. (1994). *Nature Struct. Biol.* **1**, 439.
- Okazaki, A., Ikura, T. & Kuwajima, K. (1995). *Nature Struct. Biol.* **2**, 10.
- Osterhout, J. J. Jr., Baldwin, R. L., York, E. J., Stewart, J. M., Dyson, H. T. & Wright, P. E. (1989). *Biochemistry*, **28**, 7059.
- Pain, R. H. (1994). *Mechanisms*, IRL Press, Oxford.
- Painter, P. C. & Koenig, J. L. (1976). *Biopolymers*, **15**, 229.
- Pancoska, P., Bitto, E., Janota, V. & Keiderling, T. A. (1994). *Faraday Discuss.* **99**, 287.
- Papiz, M. Z., Sawyer, L., Eliopoulos, E. E., North, A. C. T., Findlay, J. B. C., Sivaprasadarao, R., Jones, T. A., Newcomer, M. E. & Kraulis, P. J. (1986). *Nature*, **324**, 383.
- Paselk, R. A. & Levy, D. (1974). *Biochim. Biophys. Acta*, **3**, 215.
- Paterlini, M. G., Freedman, T. B. & Nafie, L. A. (1986). *Biopolymers*, **25**, 1751.

- Pedersen, T. G., Sigurskjold, B. W., Andersen, K. V., Kjaer, M., Poulsen, F. M., Dobson, C. M. & Redfield, C. (1991). *J. Mol. Biol.* **218**, 413.
- Pemberton, J. E. & Sobocinsky, R. L. (1989). *J. Am. Chem. Soc.* **111**, 432.
- Permyakov, E. A., Yarmolenko, V. V., Kalinichenko, L. P., Morozova, L. A. & Burstein, E. A. (1981). *Biochem. Biophys. Res. Commun.* **100**, 191.
- Perutz, M. (1992). *Protein Structure: New Approaches to Disease and Therapy*, W. H. Freeman and Company, New York.
- Piez, K. A. & Sherman, M. R. (1970). *Biochemistry*, **9**, 4129.
- Polavarapu, P. L. (1990). *J. Phys. Chem.* **94**, 8106.
- Polavarapu, P. L., Black, T. M., Barron, L. D. & Hecht, L. (1993a). *J. Am. Chem. Soc.* **115**, 7736.
- Polavarapu, P. L., Bose, P. K., Hecht, L. & Barron, L. D. (1993b). *J. Phys. Chem.* **97**, 11211.
- Polavarapu, P. L. & Deng, Z. (1994). *Faraday Discuss.* **99**, 151.
- Prestrelski, S. J., Blyer, D. M. & Thompson, M. P. (1991). *Biochemistry*, **30**, 8797.
- Prestrelski, S. J., Blyer, D. M. & Thompson, M. P. (1996). *Int. J. Peptide Protein Res.* **37**, 508.
- Ptitsyn, O. B. (1992). In *Protein Folding*, (Creighton, T. E. ed), pp. 243 - 300, Freedman, New York.
- Pullen, R. A., Lindsay, D. G., Wood, S. P., Tickle, I., Blundell, T. L., Wollmer, A., Krail, A., Bradenburg, I., Zahn, H., Gliemann, J. & Gammeltoft, S. (1976). *Nature (London)*, **259**, 369.
- Purcell, F. (1993). *Laser Focus World*, vol. 29, No.10, pp. 135.
- Redfield, C. & Dobson, C. M. (1988). *Biochemistry*, **27**, 122.
- Reed, L. L. & Johnson, P. L. (1973). *J. Am. Chem. Soc.* **95**, 7523.
- Reeke, G. N., Becker, J. W. & Edelman, G. M. (1975). *J. Biol. Chem.* **250**, 1525.
- Robinson, C. V., Groß, M., Eyles, S. J., Ewbank, J. J., Mayhew, M., Hartl, F. U., Dobson, C. M. & Radford, S. E. (1994). *Nature*, **372**, 646.

- Rosenfield, L. (1928). *Z. Phys.* **52**, 161.
- Santoro, J., González, C., Bruix, M., Neira, J. L., Nieto, J. L., Herranz, J. & Rico, M. (1993). *J. Mol. Biol.* **229**, 722.
- Sasisekharan, V. (1954). *Acta Cryst.* **12**, 897.
- Schmid, K., Kaufmann, H., Isemura, S., Bauer, F., Emura, J. & Motoyama, T. (1973). *Biochemistry*, **12**, 2711.
- Sengupta, P. K. & Krimm, S. (1987). *Biopolymers*, **26**, 599.
- Sepulveda, P., Marciniszyn, J., Liu, D. & Tang, J. (1975). *J. Biol. Chem.* **250**, 5082.
- Sewell, G. L. (1986). *Quantum Theory of Collective Phenomena*, Clarendon Press, Oxford.
- Shakhnovich, E. I. & Finkelstein, A. V. (1989a). *Biopolymers*, **28**, 1667.
- Shakhnovich, E. I. & Finkelstein, A. V. (1989b). *Biopolymers*, **28**, 1681.
- Shimizu, A., Ikeguchi, M. & Sugai, S. (1993). *Biochemistry*, **32**, 13198.
- Shortle, D. (1993). *Curr. Opin. Struct. Biol.* **3**, 66.
- Sielecki, A. R., Fedorov, A. A., Boodhoo, A., Andreeva, N. S. & James, M. N. G. (1990). *J. Mol. Biol.* **214**, 143.
- Sigler, P. B. & Blow, D. M. (1965). *J. Mol. Biol.* **14**, 640.
- Sjöholm, I. & Ljungstedt, I. (1973). *J. Biol. Chem.* **248**, 8434.
- Smith, L. J., Sutcliffe, M. J., Redfield, C. & Dobson, C. M. (1993). *J. Mol. Biol.* **29**, 930.
- Smythe, H. L., Huston, S. E. & Marshall, G. R. (1995). *J. Am. Chem. Soc.* **117**, 5445.
- So, L. L. & Goldstein, I. J. (1968). *Biochim. Biophys. Acta*, **165**, 398.
- Spencer, K. M., Freedman, T. B. & Nafie, L. A. (1988). *Chem. Phys. Lett.* **149**, 367.
- Stein, D. L. (1985). *Proc. Natl. Acad. Sci. USA*, **82**, 3670.
- Stein, D. L. (1992). *Spin Glasses and Biology*, World Scientific, Singapore.

- Stein, P. E., Leslie, A. G. W., Finch, J. T. & Carrell, R. W. (1991). *J. Mol. Biol.* **221**, 941.
- Stephens, P. J. & Lowe, M. A. (1985). *Annu. Rev. Phys. Chem.* **36**, 213.
- Stockman, B. J., Nirmala, N. R., Wagner, G., Delcamp, T. J., De Yarman, M. T. & Freisheim, J. H. (1992). *Biochemistry*, **31**, 218.
- Sugai, S. & Ikeguchi, M. (1994). *Advan. Biophys.* **30**, 37.
- Sugana, K. R. R., Bolt, E. A., Padlan, E., Subramanian, S., Sheriff, G. H., Cohen, A. & Davies, D. R. (1987). *J. Mol. Biol.* **196**, 877.
- Swindells, M. B., Mac Arthur, M. W. & Thornton, J. M. (1995). *Nature Struct. Biol.* **2**, 596.
- Takada, K., Chida, K. & Noda, J. (1988). *J. Opt. Soc. Am.* **A11**, 1905.
- Tedesco, J. M., Owen, H., Pallister, D. M. & Morris, M. D. (1993). *Anal. Chem.* **65**, 441A.
- Tiffany, M. L. & Krimm, S. (1968a). *Biopolymers*, **6**, 1379.
- Tiffany, M. L. & Krimm, S. (1968b). *Biopolymers*, **6**, 1767.
- Tiffany, M. L. & Krimm, S. (1969). *Biopolymers*, **8**, 347.
- Traub, W. & Shmueli, U. (1963). *Nature*, **198**, 1165.
- Tsukada, H. & Blow, D. M. (1985). *J. Mol. Biol.* **184**, 703.
- Tu, A. T. (1986). In *Advances in Spectroscopy* (Clark, R. J. H. & Hester, R. E. eds), vol. 13, pp. 47 - 112, Wiley, Chichester.
- Turner, J. J., Gordon, C. M. & Howdle S. M. (1995). *J. Phys. Chem.* **99**, 17532.
- Ueda, T., Nakashima, A., Hashimoto, Y., Miki, T., Yamada, H. & Imoto, T. (1994). *J. Mol. Biol.* **235**, 1312.
- Urbanova, M., Dukor, R. K., Pancoska, P., Gupta, V. P. & Keiderling, T. A. (1991). *Biochemistry*, **30**, 10479.
- Wang, J. L., Cunningham, B. A. & Edelman, G. M. (1971). *Proc. Nat. Acad. Sci. USA*, **68**, 1130.

- Weisgerber, S. & Helliwell, J. R. (1993). *J. Chem. Soc. Faraday Trans.* **89**, 2667.
- Wen, Z. Q. (1992). Doctoral Thesis, Glasgow University, 1992.
- Wen, Z. Q., Hecht, L. & Barron, L. D. (1994a). *J. Am. Chem. Soc.* **116**, 443.
- Wen, Z. Q., Hecht, L. & Barron, L. D. (1994b). *Protein Sci.* **3**, 435.
- Williamson, J. M., Bowling, R. J. & Mc Creery, R. L. (1989). *Appl. Spectrosc.* **43**, 372.
- Williamson, K. L. & Williams, R. J. P. (1979). *Biochemistry*, **18**, 5966.
- Wilson, G., Ford, S. J., Cooper, A. Hecht, L. Wen, Z. Q. & Barron, L. D. (1995). *J. Mol. Biol.* **254**, 747.
- Wilson, G., Hecht, L. & Barron, L. D. (1996). *J. Chem. Soc. Faraday Trans.*, **92**, 1503.
- Wilson, G., Hecht, L. & Barron, L. D. (1996). *J. Mol. Biol.* in press.
- Wlodawer, A. & Sjölin, L. (1983). *Biochemistry*, **22**, 2720.
- Wlodawer, A., Borkakoti, N., Moss, D. S. & Howlin, B. (1986). *Acta Cryst. Sect. B*, **42**, 379.
- Wlodawer, A., Svensson, L. A., Sjölin, L. & Gilliland, G. L. (1988). *Biochemistry*, **27**, 2705.
- Wlodawer, A. & Savage, H. (1989). *Acta Cryst.* **B45**, 99.
- Woody, R. W. (1977). *J. Polym. Sci. Macromol. Rev.* **15**, 229.
- Woody, R. W. (1992). *Adv. Biophys. Chem.* **2**, 37.
- Woody, R. W. (1994). In *Circular Dichroism Principles and Applications*, (Nakanishi, K., Berova, N. & Woody, R. W. eds), pp. 473 - 496, VCM Publishers, New York.
- Wu, L. C., Peng, Z. & Kim, P. S. (1995). *Nature Struct. Biol.* **2**, 281.
- Yang, B., Owen, H., Pallister, D. M. & Morris, M. D. (1991). *Appl. Spectrosc.* **45**, 1533.
- Yang, J. J., Buck, M., Pitkeathly, M., Kotik, M., Haynie, D. T., Dobson, C. M. &

- Radford, S. E. (1995). *J. Mol. Biol.* **252**, 483.
- Yarin, J., Kalb, A. J. & Levitzki, A. (1968). *Biochim. Biophys. Acta*, **165**, 303.
- Yasui, S. C., Keiderling, T. A., Bonora, G. M. & Toniolo, C. (1986a). *Biopolymers*, **25**, 79.
- Yasui, S. C., Keiderling, T. A., Formaggio, F., Bonora, G. M. & Toniolo, C. (1986b). *J. Am. Chem. Soc.* **108**, 4988.
- Yoder, G., Keiderling, T. A., Formaggio, F., Crisma, M., Toniolo, C. & Kamphuis, J. (1995). *Tetrahedron: Asymmetry*, **6**, 687.
- Yu, N. -T., Liu, C. S. & O' Shea, D. C. (1972). *J. Mol. Biol.* **70**, 117.
- Yu, N. -T. & Liu, C. S. (1972). *J. Am. Chem. Soc.* **94**, 3250.
- Yu, N. -T. (1974). *J. Am. Chem. Soc.* **96**, 4664.
- Yu, T. J., Lippert, J. L. & Peticolas, W. L. (1973). *Biopolymers*, **12**, 2161.
- Yu, T. J. (1977). *CRC Critical Reviews in Biochemistry*, **4**, 229.
- Yukhnovskii, I. R. (1987). *Phase Transitions of the Second Order-Collective Variables Method*, World Scientific, Singapore.

## Publications

1. S. J. Ford, A. Cooper, L. Hecht, **G. Wilson** and L. D. Barron. *Journal of the Chemical Society Faraday Transactions*, **91** (1995) 2087 - 2093.  
"Vibrational Raman Optical Activity of Lysozyme: Hydrogen-Deuterium Exchange, Unfolding and Ligand Binding."
2. A. F. Bell, S. J. Ford, L. Hecht, **G. Wilson** and L. D. Barron. *International Journal of Biological Macromolecules*, **16** (1994) 277 - 278.  
"Vibrational Raman Optical Activity of Glycoproteins."
3. L. D. Barron, L. Hecht, S. J. Ford and **G. Wilson**. *Journal of Molecular Structure*, **349** (1995) 397 - 400.  
"Vibrational Raman Optical Activity of Biopolymers."
4. **G. Wilson**, S. J. Ford, A. Cooper, L. Hecht, Z. Wen and L. D. Barron. *Journal of Molecular Biology*, **254** (1995) 747 - 760.  
"Vibrational Raman Optical Activity of  $\alpha$ -Lactalbumin: Comparison with Lysozyme and Evidence for Native Tertiary Folds in Molten Globule States."
5. **G. Wilson**, L. Hecht and L. D. Barron. *Journal of the Chemical Society Faraday Transactions*, **92** (1996) 1503 - 1509.  
"Vibrational Raman Optical Activity of  $\alpha$ -Helical and Unordered Poly-L-Lysine."
6. L. D. Barron, L. Hecht, A. F. Bell and **G. Wilson**. *Applied Spectroscopy*, in press.  
"Vibrational Raman Optical Activity: From Fundamentals to Biochemical Applications."
7. **G. Wilson**, L. Hecht and L. D. Barron. *Journal of Molecular Biology*, in press.  
"The Native-Like Tertiary Fold in Molten Globule  $\alpha$ -Lactalbumin Appears to be Controlled by a Continuous Phase Transition."

# Vibrational Raman Optical Activity of Lysozyme: Hydrogen-Deuterium Exchange, Unfolding and Ligand Binding

Steven J. Ford, Alan Cooper, Lutz Hecht, Gary Wilson and Laurence D. Barron\*  
Chemistry Department, The University, Glasgow, UK G12 8QQ

Measurements of the vibrational Raman optical activity (ROA) spectra of hen egg white lysozyme are reported which show that ROA is a useful new probe of protein secondary and tertiary structure and dynamics. ROA spectra can be measured just as easily in  $D_2O$  as in  $H_2O$  and a comparison of the two gives information about the relative exchange rates of the amide hydrogens in the peptide backbone for the various types of secondary and tertiary structure in lysozyme. Unfolded lysozyme shows a large conservative ROA couplet in the amide III region which might facilitate the identification of signatures in the ROA spectra of native proteins from irregular structures with the same type of conformational heterogeneity as that of an unfolded protein. The ROA spectrum of lysozyme bound to a saccharide inhibitor shows evidence for an increase in rigid loop content.

Determination of the solution structure of proteins remains a central problem in protein science. Vibrational spectroscopy, both IR and Raman, has been widely used in such studies but provides only rather coarse conformational information.<sup>1-4</sup> However, the potential value of vibrational spectroscopy in this area has now been enhanced by adding the new dimension of optical activity,<sup>5-7</sup> which confers an exquisite sensitivity to molecular conformation. The vibrational circular dichroism (VCD) approach, which measures a small difference in absorbance of left- and right-circularly polarized IR radiation by chiral molecules, is now well established as a valuable probe of protein secondary structure.<sup>8-10</sup> Until recently, however, lack of instrument sensitivity has restricted application of the complementary technique of vibrational Raman optical activity (ROA), which measures a small difference in the intensity of vibrational Raman scattering from chiral molecules in right- and left-circularly polarized incident light, to more favourable samples.<sup>11</sup> However, thanks to new developments in instrumentation,<sup>12,13</sup> ROA can now be measured routinely on a large range of biological molecules in aqueous solution and appears to provide a completely new perspective on their solution structure.<sup>14-17</sup> Proteins show particularly rich ROA spectra which, in addition to signatures characteristic of secondary backbone and sidegroup conformation, also appear to contain signatures from loops and turns so that supersecondary structural elements such as helix-loop-helix motifs might be discerned from a complete protein ROA spectrum.<sup>18,19</sup> Although it is unlikely that ROA will provide complete structures like X-ray crystallography and multi-dimensional NMR, its simple application to aqueous solution samples with no restrictions on the size of the biopolymer (unlike multi-dimensional NMR) makes it attractive for studying many current problems in protein chemistry.

The normal vibrational modes of proteins can be highly complex, with contributions from many skeletal and side group local vibrational coordinates, all of which can affect the intensity and frequency. However, the great virtue of ROA is its ability to cut through some of the inherent complexity of conventional vibrational spectra: only those few local vibrational coordinates within a complicated normal mode which sample the skeletal chirality most directly make significant contributions to the associated ROA intensity, thereby generating characteristic ROA band patterns which are usually much simpler and more informative than the parent Raman band patterns.

Unlike VCD, which in most cases has to rely on  $D_2O$  solutions for clear signal identification, ROA measurements can

be performed as easily in  $H_2O$  as in  $D_2O$ . A central purpose of the present work was to see how exchange of the amide hydrogen atoms for deuterium atoms in  $D_2O$  can assist in the assignment of protein ROA bands, and to investigate how ROA might be used to monitor the accessibility of amide hydrogen atoms to deuterium exchange in different types of secondary and tertiary structure.

Hen egg white lysozyme is an anti-bacterial enzyme with a  $\beta(1-4)$  glucoaminidase activity which allows it to catalyse hydrolysis of the walls of bacterial cells. It belongs to the c-type lysozyme family which encompasses all the mammalian lysozymes, moth lysozymes and some bird lysozymes. The molecule contains 129 amino acid residues and has a molecular weight of 14.3 kD. The structure as deduced from X-ray crystallography<sup>20</sup> is shown schematically in Fig. 1 (drawn using the program MolScript<sup>21</sup>). There are four  $\alpha$ -helices which make up one domain encompassing the amino- and carboxy-terminal segments, and a triple-stranded antiparallel  $\beta$ -sheet that, together with a long loop, makes up much of a second domain. In addition there are two  $3_{10}$  helices, one in each domain, and a short region of double-stranded antiparallel  $\beta$ -sheet linking the two domains. The structure is stabilized by four disulfide bridges (6  $\leftrightarrow$  127, 30  $\leftrightarrow$  115, 64  $\leftrightarrow$  80 and 76  $\leftrightarrow$  94). The locations of the secondary elements, as taken from ref. 25, are listed in Table 1.

The aqueous solution structure of hen egg white lysozyme has been determined using two-dimensional (2D)  $^1H$  NMR spectroscopy.<sup>22,23</sup> The secondary structure and main chain fold were found to be virtually identical to that in the crystal; however, this was not the case for the surface residues. Although lacking the tremendous structural detail provided

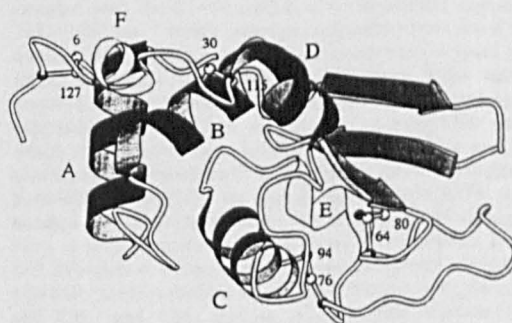


Fig. 1 Schematic structure of hen egg white lysozyme<sup>20</sup>



**Table 1** Secondary structure elements in hen egg white lysozyme<sup>25</sup>

structure	residues
$\alpha$ -helix	4-15 (A), 24-36 (B), 88-99 (C), 108-115 (D)
$3_{10}$ -helix	79-84 (E), 119-124 (F)
$\beta$ -sheet	1-2, 39-40 (double-stranded)
	42-46, 50-54, 57-60 (triple-stranded)

by the recent NMR studies, it is worth mentioning that an early study<sup>24</sup> which compared the conventional Raman spectra of hen egg white lysozyme in the crystal and in solution also concluded that the main chain conformations were very similar.

We chose hen egg white lysozyme for these studies because it is readily available, is well characterized by numerous X-ray, NMR and other studies,<sup>25</sup> and is sufficiently robust to withstand the procedures necessary to ensure complete hydrogen exchange. It also served as a useful model for the parallel ROA studies reported here of the effects of unfolding and ligand binding on the solution structure of proteins.

### Experimental

The instrument used for the ROA measurements has been described previously.<sup>12</sup> It is based on backscattering, which is essential for aqueous solution samples such as proteins, and employs a single-grating spectrograph fitted with a back-thinned CCD camera as detector and a holographic notch filter to block the Rayleigh line. ROA is measured by synchronizing the spectral acquisition with an electro-optic modulator used to switch the polarization of the incident argon-ion laser beam between right and left circular at a suitable rate.

We used Sigma Grade I hen egg white lysozyme. All the H<sub>2</sub>O buffer solutions were prepared using Analar grade chemicals and distilled and deionized water. The D<sub>2</sub>O used for the deuteriated buffers was purchased from Fluorochem. The pH values quoted for the deuteriated solutions are uncorrected but can be converted to pD values by adding 0.41.<sup>26</sup> The protein solutions were prepared in small glass sample tubes and shaken with a small amount of pharmaceutical grade activated charcoal, subsequently removed by centrifugation, to remove traces of fluorescing impurities. Typical protein concentrations were 80 mg ml<sup>-1</sup>. The solutions were filtered through Millipore GV4 (0.22  $\mu$ m) filters into quartz microfluorescence cells which were subsequently centrifuged. Residual fluorescence was allowed to 'burn down' by leaving the sample in the laser beam for several hours before acquiring the ROA spectra.

Denatured (or, more precisely, unfolded) lysozyme was prepared by reducing all four disulfide bridges using the method of Creighton,<sup>27</sup> modified at the purification stage to dialysis against six 3 l volumes of 0.01 mol l<sup>-1</sup> HCl. We used this method of denaturation because the other commonly-used methods, namely chemical denaturation and thermal denaturation, provide samples which are unsuitable for ROA studies (chemically-denatured proteins have intense Raman bands from the denaturants and thermally-denatured proteins often show considerable light scatter owing to coagulation). Disulfide reduction is reversible so the free sulfhydryl groups are usually carboxymethylated to prevent re-oxidation; however, we found that the resulting carboxymethylated protein was insufficiently soluble for ROA studies, so we left the sulfhydryl groups unblocked and maintained a low pH to prevent re-oxidation as was done in recent NMR studies of denatured lysozyme.<sup>28</sup> The number of

free sulfhydryl groups was checked using the Ellman assay<sup>27</sup> and found to be between seven and eight per protein molecule. The solution was lyophilized and finally prepared for ROA measurements by dissolving in citrate buffer at pH 3.0.

Partially deuteriated lysozyme samples were made up by dissolving the native protein in deuteriated acetate buffer at 4.0 and leaving the solutions for either seven or 25 days at room temperature. They were then refrigerated prior to preparation for ROA analysis. Completely deuteriated samples in which all the amide hydrogen atoms have been replaced by deuterium atoms can only be prepared by denaturing the protein reversibly in the presence of D<sub>2</sub>O or a deuteriated buffer. We employed thermal denaturation: the lysozyme was dissolved in D<sub>2</sub>O and left in an oven at 60°C for five days, following which it was lyophilized and then prepared as a normal lysozyme sample by dissolving in deuteriated acetate buffer at pH 4.0. Tests on this completely deuteriated lysozyme showed that the enzymatic activity, as measured by the standard Sigma procedure of cell wall hydrolysis of *Micrococcus lysodeikticus*, was fully retained.

The preparation of the complex of lysozyme with the trimer of *N*-acetylglucosamine (NAG<sub>3</sub>) for ligand binding studies simply involved mixing the protein in acetate buffer with a two molar excess of NAG<sub>3</sub> (an excess of at least 1.3 molar is required to ensure that the protein molecules are fully bound under these conditions<sup>29</sup>). Since NAG<sub>3</sub> is slowly hydrolysed by hen egg white lysozyme, the ROA spectra were acquired within a few hours of preparing the complex (from the data in ref. 30, we deduced that the half-life of NAG<sub>3</sub> under our experimental conditions is ca. 6 days).

The ROA spectra are presented in the form of a circular intensity difference (in analogue-to-digital converter counts)  $I^R - I^L$ , where  $I^R$  and  $I^L$  are the Raman-scattered intensities in right- and left-circularly polarized incident light. The conventional Raman intensities are presented as a corresponding circular intensity sum  $I^R + I^L$ . The experimental conditions were as follows: laser wavelength, 514.5 nm; laser power at the sample, 700 mW; spectral band width, 10 cm<sup>-1</sup>; recording time, ca. 10–20 h depending on the sample concentration.

### Results and Discussion

#### General Features of the Native Lysozyme ROA Spectrum

The backscattered Raman and ROA spectra of native lysozyme in acetate buffer (H<sub>2</sub>O) at pH = 4.5 are shown in Fig. 2. The ROA spectrum is very similar to that obtained previously in unbuffered H<sub>2</sub>O.<sup>14,31</sup> Vibrations of the peptide backbone in proteins are usually associated with four main regions of the Raman spectrum:<sup>3,32</sup> the C<sub>α</sub>–C stretch region ca. 870–950 cm<sup>-1</sup>, the C<sub>α</sub>–N stretch region ca. 1020–1150 cm<sup>-1</sup>, the amide III region ca. 1230–1310 cm<sup>-1</sup> which is supposed to involve mainly the in-plane N–H deformation coupled with the C<sub>α</sub>–N stretch, and the amide I region ca. 1645–1680 cm<sup>-1</sup> which arises predominantly from the C=O stretch. In fact Diem *et al.*<sup>6,33</sup> have shown that, in small peptides, the amide III region involves much more mixing between the N–H and C<sub>α</sub>–H deformations than previously supposed, and should be extended to at least 1340 cm<sup>-1</sup>, which is supported by a detailed ROA study of a series of alanyl peptide oligomers.<sup>34</sup> This extended amide III region is particularly important for peptide and protein ROA work because the coupling between N–H and C<sub>α</sub>–H deformations is very sensitive to the geometry and generates a rich and informative ROA band structure. Sidegroup vibrations, especially those involving tryptophan, tyrosine, phenylalanine and CH<sub>2</sub> and CH<sub>3</sub> groups, also generate characteristic Raman bands some of which show ROA signals.

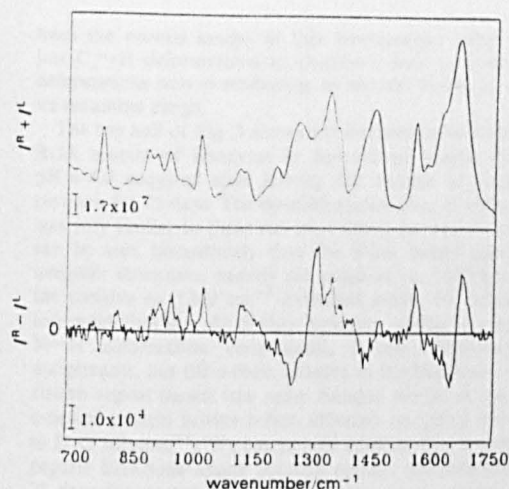


Fig. 2 Backscattered Raman ( $I^R + I^L$ ) and ROA ( $I^R - I^L$ ) spectra of native lysozyme in acetate buffer at pH 4.5

The backbone skeletal stretch region has been shown to contain ROA bands originating in  $\alpha$ -helix<sup>18</sup> and  $\beta$ -sheet<sup>19</sup> vibrations. Here the positive ROA intensity in the range ca. 880–960  $\text{cm}^{-1}$  with two distinct peaks at ca. 910 and 940  $\text{cm}^{-1}$  can be assigned to  $\alpha$ -helix modes, as can the positive ROA band at ca. 1120  $\text{cm}^{-1}$  and perhaps also the associated small negative ROA band at ca. 1095  $\text{cm}^{-1}$ . The positive ROA band at ca. 1020  $\text{cm}^{-1}$  might originate in  $\beta$ -sheet modes, but at present we cannot rule out the possibility that aromatic sidegroups are involved.

Unlike VCD in the amide I region, which is very sensitive to different types of protein secondary structure,<sup>8–10</sup> ROA in the amide I region is not particularly informative. Both  $\alpha$ -helix and  $\beta$ -sheet appear to give amide I couplets, negative at low wavenumber and positive at high; this couplet is usually conservative in  $\beta$ -sheet proteins,<sup>19</sup> but has more intensity at higher wavenumber in  $\alpha$ -helix proteins.<sup>18</sup> The appearance of the amide I lysozyme ROA couplet in Fig. 2 is consistent with a superposition of  $\alpha$ -helix and  $\beta$ -sheet couplets since the positive higher-wavenumber component is a little more intense than the negative lower-wavenumber component. However, the protein amide I Raman band is usually intense and quite strongly polarized which can lead to distortions in the associated ROA; also the contributions from irregular structure in this region have not been fully disentangled (the ROA spectrum of unfolded lysozyme discussed below sheds some light on this). Hence further studies are necessary before these particular signatures can be applied with confidence.

The most interesting region for protein ROA studies appears to be the extended amide III. Although the overall appearance of the ROA spectrum here is similar in all proteins studied to date, namely a broad couplet negative at low wavenumber and positive at high, there is considerable variation in the detailed structure which seems highly informative. The lysozyme ROA spectrum in Fig. 2 shows a strong sharp positive peak at ca. 1300  $\text{cm}^{-1}$ , with a negative shoulder at ca. 1270  $\text{cm}^{-1}$  which might be characteristic of  $\alpha$ -helix,<sup>18</sup> but there is no evidence of the sharp positive band at ca. 1313  $\text{cm}^{-1}$  shown by  $\beta$ -sheet proteins,<sup>19</sup> presumably because the low  $\beta$ -sheet content of lysozyme generates only small ROA intensity here which is swamped by the intense band at ca. 1300  $\text{cm}^{-1}$ . The strong sharp negative ROA band at ca. 1245  $\text{cm}^{-1}$  and the sharp positive ROA band at ca. 1340  $\text{cm}^{-1}$  do

not arise from  $\alpha$ -helix or  $\beta$ -sheet vibrations, and we have previously ascribed them to two types of protein loop structure: since insulin and bovine serum albumin, respectively, show particularly clear examples of the lower and higher wavenumber bands, we have said that they are signatures of 'insulin-type' loops and 'BSA-type' loops.<sup>18</sup> The assignment of these two ROA bands to loop (or at least irregular) structures is reinforced by the unfolding and deuteration studies discussed later.

Sidegroup bands can also be discerned in the lysozyme ROA spectrum (Fig. 2). Lysozyme contains five tryptophan residues which are responsible, among other things, for the strong positive ROA associated with the strong Raman band at ca. 1555  $\text{cm}^{-1}$  and the weak negative ROA associated with the weaker Raman band at ca. 1580  $\text{cm}^{-1}$ . Although tryptophan modes are assigned to the Raman peaks at ca. 1336 and 1355  $\text{cm}^{-1}$  which constitute a Fermi resonance doublet,<sup>35</sup> there are three pieces of evidence which suggest that the ROA observed in this region originates mainly in extended amide III modes rather than in tryptophan modes.<sup>36</sup> First, a very strong positive ca. 1340  $\text{cm}^{-1}$  ROA band appears in bovine serum albumin<sup>18</sup> which contains only two tryptophans but there are no ROA signals in any of the other regions where tryptophan signals appear; secondly, the deuterium exchange experiments (*vide infra*) show that when most of the tryptophan NH protons should be exchanged a small positive ca. 1340  $\text{cm}^{-1}$  ROA signal can still be seen; and thirdly, in the unfolded protein (*vide infra*) the tryptophan ROA signal at ca. 1555  $\text{cm}^{-1}$  can still be clearly seen whereas the ca. 1340  $\text{cm}^{-1}$  ROA signal vanishes completely. The shoulder at ca. 1620  $\text{cm}^{-1}$  on the amide I Raman band originates in aromatic modes<sup>32,35</sup> and is associated with a small negative ROA signal. The strong Raman band at ca. 1451  $\text{cm}^{-1}$  with a shoulder at 1430  $\text{cm}^{-1}$  originates in sidegroup  $\text{CH}_2$  and  $\text{CH}_3$  deformations, with a contribution from tryptophan (mostly the NH bend of the indole ring) to the ca. 1430  $\text{cm}^{-1}$  shoulder,<sup>35,37</sup> and is associated with an ROA couplet negative at low wavenumber and positive at high. Little information has yet been extracted from sidegroup ROA features such as these.

### Hydrogen–Deuterium Exchange

It has been recognized for some time that hydrogen–deuterium exchange methods can be of great value in the study of the dynamic aspects of protein structure.<sup>38</sup> However, the full power of this approach was only unleashed after 2D NMR techniques became available for measuring exchange rates for a large number of individual backbone and side-chain amide protons in a protein. For a native protein dissolved in  $\text{D}_2\text{O}$  this can provide information about the relative accessibility of different types of structure to the solvent: the rate of exchange is decreased by many orders of magnitude when the NH groups are involved in hydrogen bonds within secondary structures such as helices and sheets or in some tertiary interactions, whereas NH groups in surface loop and end chain structures usually exchange rapidly. Lysozyme has been studied extensively in this way.<sup>39,40</sup>

Hydrogen–deuterium exchange is also much used in vibrational spectroscopy as a valuable aid to band assignment. Because some NH groups exchange very rapidly and others very slowly, a comparison of the ROA spectrum of a protein in  $\text{H}_2\text{O}$  and  $\text{D}_2\text{O}$  solutions can give information about dynamics as well as band assignment (indeed, band assignments can depend on exchange rates). The extended amide III region is particularly informative because replacement of N–H by N–D removes the crucial N–H deformations

from the normal modes in this wavenumber range leaving just  $C_\alpha-H$  deformations to dominate here with the  $N-D$  deformations now contributing to normal modes in a lower wavenumber range.

The top half of Fig. 3 shows the backscattered Raman and ROA spectra of lysozyme in deuterated acetate buffer at pH = 4.0 acquired after leaving the sample at room temperature for 25 days. The detailed appearance of both spectra was very similar to those run after seven days (not shown). It can be seen immediately that the ROA bands ascribed to irregular structures, namely the negative *ca.* 1245  $cm^{-1}$  and the positive *ca.* 1340  $cm^{-1}$  extended amide III bands, have lost much (but not all) of their intensity, owing to loss of the  $N-H$  deformation coordinates, thereby reinforcing the assignments, but the  $\alpha$ -helix features in the backbone skeletal stretch region persist (the same features persist in the highly  $\alpha$ -helical protein bovine serum albumin on going from  $H_2O$  to  $D_2O$  solution<sup>17</sup>). We can indeed confirm that a number of peptide backbone amide protons remain unexchanged after 25 days for, as can be seen from the spectra shown in the bottom half of Fig. 3, further significant changes in the ROA spectrum occur after subjecting the sample to the rigorous procedure, described in the experimental section above, which ensures complete  $H-D$  exchange.

It is useful at this point to review the details of the deuterium exchange behaviour of lysozyme provided by NMR studies.<sup>39,40</sup> Pedersen *et al.*<sup>39</sup> divided the amide groups of lysozyme into the following four categories depending on their rate of deuterium exchange following dissolution of the

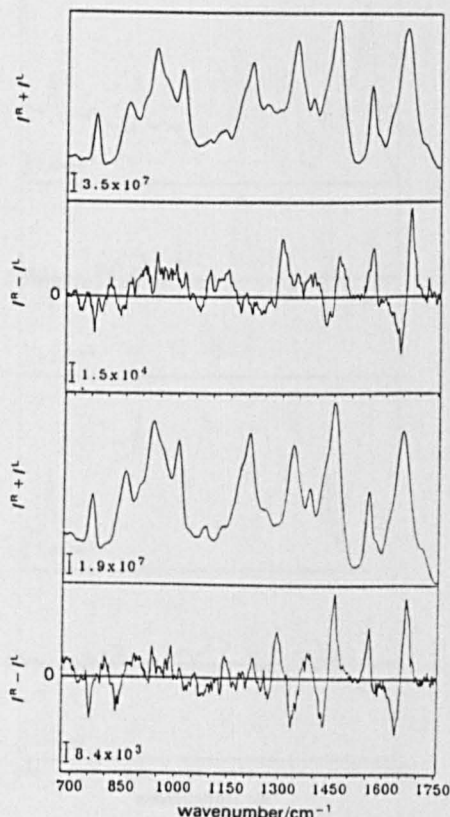


Fig. 3 Backscattered Raman and ROA spectra of lysozyme in deuterated acetate buffer at pH 4.0 after 25 days (top half), and after complete exchange of the amide protons (bottom half)

protein in  $D_2O$  at 22°C: categories I and II with greater than 90% exchange at pH = 4.2 after 12 h and 25 days, respectively, and categories III and IV with greater than and less than 90% exchange, respectively, at pH = 7.5 after 25 days.

Our '25-day' sample would correspond approximately to exchange of category I and II amides, the positions of which are shown in Fig. 4 (according to ref. 39). The ROA spectrum of the 25-day sample is consistent with these exchange data which show that most of the non-helix and non-sheet structure has exchanged. The residual intensity seen in the *ca.* 1340  $cm^{-1}$  positive ROA band therefore presumably originates in the few unexchanged amides in the relevant loop sections.

The largest changes between the completely undeuterated and the fully deuterated lysozyme ROA spectra occur in the region *ca.* 1200–1500  $cm^{-1}$ . Apart from an overall shift of *ca.* 25–30  $cm^{-1}$  to lower wavenumber, it is striking how similar the ROA band pattern of the fully deuterated protein is to that of small L-alanyl peptide oligomers in  $D_2O$  in this region.<sup>34</sup> Thus, as in the alanyl peptides, we can attribute the ROA couplet with a positive peak at *ca.* 1290  $cm^{-1}$  and a negative peak at *ca.* 1333  $cm^{-1}$  to coupled orthogonal  $C_\alpha-H$  deformations generated by the local chiral environments of each  $\alpha$ -carbon atom, and two pieces of evidence suggest that it originates at least in part in the fully deuterated  $\alpha$ -helix structures: first, a similar ROA couplet is shown by deuterated  $\alpha$ -helical poly-L-lysine;<sup>41</sup> secondly, the ROA spectrum of unfolded lysozyme in  $D_2O$  (*vide infra*) does not show this feature. However, both random coil poly-L-lysine and poly-L-glutamic acid in  $D_2O$  show a similar feature<sup>41</sup> so it is probably not a unique signature of the deuterated  $\alpha$ -helix. The new band appearing in the Raman spectrum of the deuterated protein at *ca.* 1386  $cm^{-1}$ , associated with a large positive ROA, originates in the tryptophan  $N-H$  bending mode identified at *ca.* 1430  $cm^{-1}$  in the undeuterated sample (*vide supra*) which has shifted to lower wavenumber on replacing  $N-H$  by  $N-D$ . This band is weaker in the 25-day sample than in the fully deuterated sample because some tryptophans are less exposed than others to the solvent, as quantified from NMR measurements of tryptophan exchange rates<sup>42</sup> (in fact the *ca.* 1386  $cm^{-1}$  Raman band has actually been used to follow the kinetics of tryptophan exchange<sup>43</sup>). The small ROA couplet, negative at *ca.* 1425  $cm^{-1}$  and positive at *ca.* 1470  $cm^{-1}$ , in the undeuterated protein has grown to dominate the ROA spectrum of the completely deuterated protein: as in the alanyl peptides,<sup>34</sup> we can attribute the high-wavenumber component to amide II' modes which have shifted down from the amide II region at *ca.* 1570  $cm^{-1}$  on deuteriation so that the associated in-plane  $N-D$  deformation now mixes with the sidegroup  $CH_2$  and  $CH_3$  deformations.

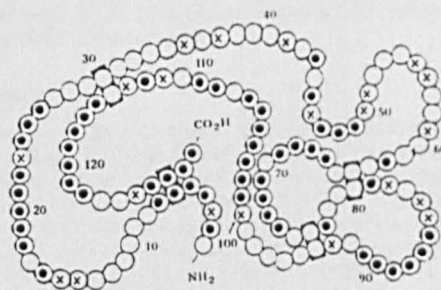


Fig. 4 Positions of amide hydrogens in lysozyme which would be fully exchanged with deuterium after 12 h (black dots, category I) and 25 days (crosses, category II)<sup>39</sup>



The ROA couplet in the amide I region has shifted by ca.  $9\text{ cm}^{-1}$  to lower wavenumber in the fully deuterated protein compared with the 25-day and undeuterated samples, which supports the suggestion that it arises largely from  $\alpha$ -helix and  $\beta$ -sheet structures (this shift of the amide I on deuteration is due to a small contribution of the N—H in-plane bend to the normal mode<sup>1</sup>).

### ROA of Unfolded Lysozyme

The backscattered Raman and ROA spectra of unfolded lysozyme (prepared by reducing all four disulfide bridges as described above) in citrate buffer at pH = 3 is shown in the top half of Fig. 5. Some differences are apparent in the conventional Raman spectrum compared with that of the native protein in Fig. 2: these were studied in detail in the early days of conventional protein Raman spectroscopy and were attributed to changes in the environment of the sidegroups together with loss of secondary structure in the peptide backbone, although all that could be said about the latter was that it adopted a 'random coil' conformation.<sup>44,45</sup>

The ROA spectrum of unfolded lysozyme has several interesting features. It is immediately apparent that much of the ROA structure has disappeared, with the spectrum now being dominated by a large, featureless conservative couplet in the amide III region. A small amide I couplet with more intensity in the higher wavenumber component remains, but is shifted along with its parent Raman band to higher wavenumber by

ca.  $10\text{ cm}^{-1}$  in accord with the amide I assignment of 'random coil' in conventional Raman spectroscopy of proteins.<sup>3,44,45</sup> This residual amide I couplet presumably arises from the local chiral environments of the carbonyl oscillators. A few weak sidegroup features can still be discerned: in particular, the positive ROA in the ca.  $1550\text{ cm}^{-1}$  tryptophan band is still clear although reduced in intensity; and the negative ROA in the ca.  $1620\text{ cm}^{-1}$  aromatic band is now clearly resolved from the negative part of the shifted amide I couplet.

The most important aspect of the unfolded lysozyme ROA spectrum is the insight it provides into the ROA loop signatures in the extended amide III region. This region now shows a large unstructured couplet with the negative low wavenumber component peaking at ca.  $1235\text{ cm}^{-1}$  and the positive high wavenumber component at ca.  $1300\text{ cm}^{-1}$  which, on account of the very short timescale of Raman scattering (ca.  $10^{-13}\text{ s}$ ), presumably originates in a superposition of the separate ROA spectra of each of the enormous number of distinct conformers present in the unfolded state. There is no trace of the positive 'BSA-type' ROA loop signature shown by the native protein at ca.  $1340\text{ cm}^{-1}$ , which further reinforces its association with tertiary structure. These observations suggest that stretches of loop and end-chain structure in native lysozyme which have the same high degree of conformational heterogeneity and associated flexibility as that of the unfolded protein are responsible for the broad ROA couplet which can be seen to underlie the amide III region in the ROA spectrum of native lysozyme shown in Fig. 2. The negative ca.  $1245\text{ cm}^{-1}$  ROA band in native lysozyme appears to be separate from this underlying couplet (it is sharper and peaks at slightly higher wavenumber than the negative ca.  $1235\text{ cm}^{-1}$  band in unfolded lysozyme) which suggests that this feature, together with the positive ca.  $1340\text{ cm}^{-1}$  band, originate in distinct types of loop and end-chain structure which is relatively rigid and homogeneous. It is worth mentioning that in the conventional Raman spectra of proteins the amide III band from 'random coil' structure is assigned to the ca.  $1240\text{--}1245\text{ cm}^{-1}$  region.<sup>3,44-46</sup>

The bottom half of Fig. 5 shows the backscattered Raman and ROA spectra of unfolded lysozyme in deuterated citrate buffer. The large conservative amide III couplet shown by unfolded lysozyme in  $\text{H}_2\text{O}$  has disappeared in  $\text{D}_2\text{O}$ , which is consistent with its assignment to flexible conformationally heterogeneous structures. The small negative ROA bands at ca.  $1210$  and  $1250\text{ cm}^{-1}$  and the larger positive band at ca.  $1300\text{ cm}^{-1}$  can be assigned to local  $\text{C}_\alpha\text{--H}$  deformations from the exchanged unfolded backbone structure: it is notable that these features are similar to those seen in the 25-day partially exchanged lysozyme ROA spectra (Fig. 3) where most of the loop and end-chain structure, but little of the  $\alpha$ -helix and  $\beta$ -sheet, has exchanged. Above  $1530\text{ cm}^{-1}$  the ROA spectra of unfolded lysozyme in  $\text{H}_2\text{O}$  and  $\text{D}_2\text{O}$  are very similar, as expected for the aromatic sidegroup and amide I vibrations which contribute here.

### Ligand Binding

It is interesting to see if ROA can monitor the subtle conformational changes that might take place, both in the protein and the ligand, during complex formation in biologically relevant situations such as enzyme catalysis. Lysozyme is a good sample for such studies because the X-ray crystal structure is, of course, well known, including that of the protein bound to saccharide inhibitors where small changes in the protein conformation are observed.<sup>47</sup> *N*-Acetyl glucosamine and its oligomers are typical inhibitors for lysozyme, and for the present studies the trimer  $\text{NAG}_3$  was used.

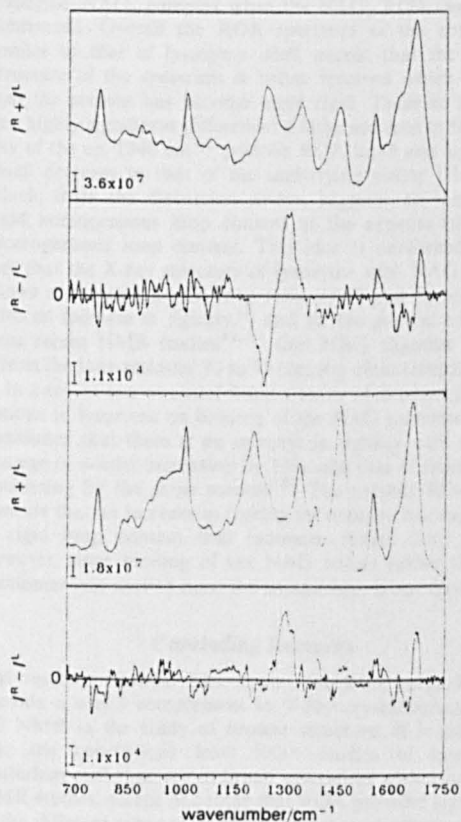


Fig. 5 Backscattered Raman and ROA spectra of unfolded lysozyme in citrate buffer (top half) and deuterated citrate buffer (bottom half) at pH 3.0

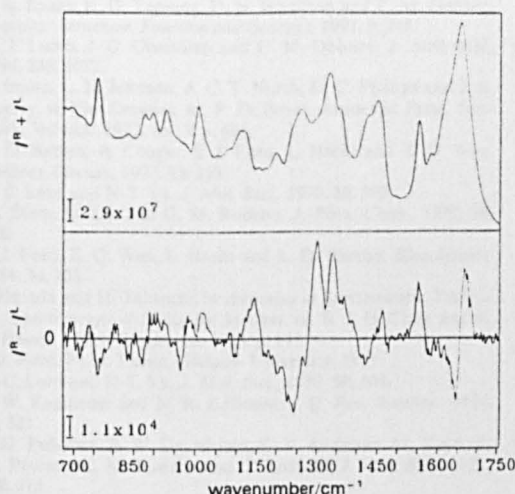


Fig. 6 Backscattered Raman and ROA spectra of lysozyme bound to NAG<sub>3</sub> in acetate buffer at pH 4.5

The backscattered Raman and ROA spectra of lysozyme bound to NAG<sub>3</sub> are shown in Fig. 6. We have also measured the Raman and ROA spectra of NAG<sub>3</sub> by itself (not shown) and, except for a reduction of the positive ROA band at ca. 1150 cm<sup>-1</sup>, find little change in the ROA spectrum of the lysozyme–NAG<sub>3</sub> complex when the NAG<sub>3</sub> ROA spectrum is subtracted. Overall the ROA spectrum of the complex is similar to that of lysozyme itself, except that the detailed structure of the spectrum is better resolved which suggests that the protein has become more rigid. There is, however, one highly significant difference: a large increase in the intensity of the ca. 1340 cm<sup>-1</sup> positive ROA band and possibly a small decrease in that of the underlying amide III couplet which, from the discussion above, suggests an increase in rigid homogeneous loop content at the expense of flexible heterogeneous loop content. This idea is reinforced by the fact that the X-ray structure of lysozyme with NAG<sub>3</sub> bound shows a change between loop residues 72 and 76 associated with an increase in rigidity,<sup>48</sup> and by the general consensus from recent NMR studies<sup>49–51</sup> that NAG oligomer binding affects the loop residues 70 to 75 (among other things).

In a recent conventional Raman study of structural modifications to lysozyme on binding of the NAG monomer it was concluded that there is an increase in rigidity with the percentage of  $\alpha$ -helix increasing by 15% and that of random coil decreasing by the same amount.<sup>52</sup> The present ROA study concurs that an increase in rigidity takes place, but suggests it is rigid loop content that increases rather than  $\alpha$ -helix; however, since binding of the NAG trimer rather than the monomer was studied here, the comparison is not direct.

### Concluding Remarks

The results presented here show that ROA measurements provide a useful complement to X-ray crystallography and 2D NMR in the study of protein structure. It is gratifying that the conclusions from ROA studies of hydrogen–deuterium exchange are in broad agreement with those from NMR studies, except of course that ROA provides signatures of the different types of gross structural units whereas NMR probes the interactions of individual residues. Our finding that ROA appears to be able to monitor loop structure and its changes on ligand binding, together with its apparent ability to discriminate between conformationally hetero-

geneous (flexible) and homogeneous (rigid) elements, is particularly interesting. It might soon be possible to specify the detailed geometry of these loop structures from *ab initio* computations of the ROA spectra of model polypeptides in a range of conformations.<sup>53</sup>

The present study, in conjunction with parallel ROA studies of the effect of temperature on lysozyme,<sup>54</sup> and of metal binding in  $\alpha$ -lactalbumin together with partial unfolding in its molten globule state,<sup>55</sup> suggests that ROA holds much promise as a new probe of the tertiary structure and dynamics of proteins in aqueous solution.

We thank the Engineering and Physical Sciences Research Council for research grants and for a Research Studentship for G.W., the Deutsche Forschungsgemeinschaft for a Research Fellowship for L.H. (Habilitationstendenzium II C1-He 1588/3-1 + 3-2), the Wellcome Trust for a Prize Studentship for S.J.F., and Dr. S. E. Radford for helpful discussions.

### References

- 1 S. Krimm and J. Bandekar, *Adv. Protein Chem.*, 1986, **38**, 181.
- 2 W. K. Surewicz and H. H. Mantsch, *Biochim. Biophys. Acta*, 1988, **952**, 115.
- 3 A. T. Tu, in *Advances in Spectroscopy, Volume 13: Spectroscopy of Biological Systems*, ed. R. J. H. Clark and R. E. Hester, Wiley Chichester, 1986, p. 47.
- 4 R. W. Williams, *Methods Enzymol.*, 1986, **130**, 311.
- 5 L. D. Barron, *Molecular Light Scattering and Optical Activity*, Cambridge University Press, Cambridge, 1982.
- 6 M. Diem, *Modern Vibrational Spectroscopy*, Wiley, New York, 1993.
- 7 L. A. Nafie, G.-S. Yu, X. Qu and T. B. Freedman, *Faraday Discuss.*, 1994, **99**, in the press.
- 8 T. A. Keiderling and P. Pancoska, in *Advances in Spectroscopy, Volume 21: Biomolecular Spectroscopy, Part B*, ed. R. J. H. Clark and R. E. Hester, Wiley, Chichester, 1993, p. 267.
- 9 T. A. Keiderling, B. Wang, M. Urbanova, P. Pancoska and R. K. Dukor, *Faraday Discuss.*, 1994, **99**, in the press.
- 10 P. Pancoska, E. Bitto, V. Janota and T. A. Keiderling, *Faraday Discuss.*, 1994, **99**, in the press.
- 11 L. D. Barron and L. Hecht, in *Circular Dichroism: Principles and Applications*, ed. K. Nakanishi, N. Berova and R. W. Woody, VCH Publishers, New York, 1994, pp. 179–215.
- 12 L. Hecht, L. D. Barron, A. R. Gargaro, Z. Q. Wen and W. Hug, *J. Raman Spectrosc.*, 1992, **28**, 401.
- 13 L. Hecht and L. D. Barron, *Faraday Discuss.*, 1994, **99**, in the press.
- 14 Z. Q. Wen, Ph.D. Thesis, Glasgow University, 1992.
- 15 L. D. Barron and L. Hecht, in *Advances in Spectroscopy, Volume 21: Biomolecular Spectroscopy, Part B*, ed. R. J. H. Clark and R. E. Hester, Wiley, Chichester, 1993, p. 235.
- 16 L. D. Barron and L. Hecht, in *Circular Dichroism: Conformational Analysis of Biomolecules*, ed. G. D. Fasman, Plenum, New York, in the press.
- 17 L. D. Barron, S. J. Ford, A. F. Bell, G. Wilson, L. Hecht and A. Cooper, *Faraday Discuss.*, 1994, **99**, in the press.
- 18 Z. Q. Wen, L. Hecht and L. D. Barron, *J. Am. Chem. Soc.*, 1994, **116**, 443.
- 19 Z. Q. Wen, L. Hecht and L. D. Barron, *Protein Science*, 1994, **3**, 435.
- 20 C. C. F. Blake, D. F. Koenig, G. A. Mair, A. C. T. North, D. C. Phillips and R. Sarma, *Nature (London)*, 1965, **206**, 757.
- 21 P. J. Kraulis, *J. Appl. Crystallogr.*, 1991, **24**, 946.
- 22 C. Redfield and C. M. Dobson, *Biochemistry*, 1988, **27**, 122.
- 23 L. J. Smith, M. J. Sutcliffe, C. Redfield and C. M. Dobson, *J. Mol. Biol.*, 1993, **229**, 930.
- 24 N.-T. Yu and B. H. Jo, *Arch. Biochem. Biophys.*, 1973, **156**, 469.
- 25 H. A. McKenzie and F. H. White, *Adv. Protein Chem.*, 1991, **41**, 173.
- 26 A. K. Covington, M. Paabo, R. A. Robinson and R. G. Bates, *Anal. Chem.*, 1968, **40**, 700.
- 27 T. E. Creighton, in *Protein Structure: a Practical Approach*, ed. T. E. Creighton, IRL Press, Oxford, 1989, p. 155.

- 28 P. A. Evans, K. D. Topping, D. N. Woolfson and C. M. Dobson, *Proteins: Structure, Function and Genetics*, 1991, 9, 248.
- 29 K. J. Lumb, J. C. Cheetham and C. M. Dobson, *J. Mol. Biol.*, 1994, 235, 1072.
- 30 T. Imoto, L. N. Johnson, A. C. T. North, D. C. Phillips and J. A. Rupley, in *The Enzymes*, ed. P. D. Boyer, Academic Press, New York, 3rd edn., 1972, vol. 7, p. 665.
- 31 L. D. Barron, A. Cooper, S. J. Ford, L. Hecht and Z. Q. Wen, *Faraday Discuss.*, 1992, 93, 259.
- 32 R. C. Lord and N-T. Yu, *J. Mol. Biol.*, 1970, 50, 509.
- 33 M. Diem, O. Lee and G. M. Roberts, *J. Phys. Chem.*, 1992, 96, 548.
- 34 S. J. Ford, Z. Q. Wen, L. Hecht and L. D. Barron, *Biopolymers*, 1994, 34, 303.
- 35 I. Harada and H. Takeuchi, in *Advances in Spectroscopy, Volume 13: Spectroscopy of Biological Systems*, ed. R. J. H. Clark and R. E. Hester, Wiley, Chichester, 1986, p. 113.
- 36 S. J. Ford, Ph.D. Thesis, Glasgow University, 1995.
- 37 R. C. Lord and N-T. Yu, *J. Mol. Biol.*, 1970, 50, 509.
- 38 S. W. Englander and N. R. Kallenbach, *Q. Rev. Biophys.*, 1984, 16, 521.
- 39 T. G. Pedersen, B. W. Sigurskjold, K. V. Andersen, M. Kjaer, F. M. Poulsen, C. M. Dobson and C. Redfield, *J. Mol. Biol.*, 1991, 218, 413.
- 40 S. E. Radford, M. Buck, K. D. Topping, C. M. Dobson and P. A. Evans, *Proteins: Struct., Funct., Genet.*, 1992, 14, 237.
- 41 G. Wilson, L. D. Barron and L. Hecht, to be published.
- 42 R. E. Wedin, M. Delepierre, C. M. Dobson and F. M. Poulsen, *Biochemistry*, 1982, 21, 1098.
- 43 H. Takesada, M. Nakanishi, A. Y. Hirakawa and M. Tsuboi, *Biopolymers*, 1976, 15, 1929.
- 44 M. C. Chen, R. C. Lord and R. Mendelsohn, *Biochim. Biophys. Acta*, 1973, 328, 252.
- 45 M. C. Chen, R. C. Lord and R. Mendelsohn, *J. Am. Chem. Soc.*, 1974, 96, 3034.
- 46 J. L. Lippert, D. Tyminski and P. J. Desmeules, *J. Am. Chem. Soc.*, 1976, 98, 7075.
- 47 C. C. F. Blake, G. A. Mair, A. C. T. North, D. C. Phillips and V. R. Sarma, *Proc. R. Soc. London, Ser. B*, 1967, 167, 365.
- 48 C. C. F. Blake, R. Cassels, C. M. Dobson, F. M. Poulsen, R. J. P. Williams and K. S. Wilson, *J. Mol. Biol.*, 1981, 147, 73.
- 49 M. Bernard, P. Canioni, P. Cozzone, J. Berthou and P. Jolles, *Int. J. Peptide Protein Chem.*, 1990, 36, 46.
- 50 I. Kumagai, K. Maenaka, F. Sunada, S. Takeda and K. Miura, *Eur. J. Biochem.*, 1993, 212, 151.
- 51 K. J. Lumb, J. C. Cheetham and C. M. Dobson, *J. Mol. Biol.*, 1994, 235, 1072.
- 52 A. Bertoluzza, S. Bonora, G. Fini and M. A. Morelli, *Can. J. Appl. Spectrosc.*, 1992, 37, 58.
- 53 P. L. Polavarapu and Z. Deng, *Faraday Discuss.*, 1994, 99, in the press.
- 54 G. Wilson, L. Hecht and L. D. Barron, to be published.
- 55 S. J. Ford, A. Cooper, L. Hecht, Z. Q. Wen and L. D. Barron, to be published.



# Preliminary Communication

## Vibrational Raman optical activity of glycoproteins

Alasdair F. Bell\*, Steven J. Ford, Lutz Hecht, Gary Wilson and Laurence D. Barron\*

Chemistry Department, The University, Glasgow G12 8QQ, UK

Received 21 June 1994; revised 4 July 1994

This paper reports the first vibrational Raman optical activity (ROA) spectrum of a glycoprotein. The sample, orosomucoid ( $\alpha_1$ -acid glycoprotein), shows ROA bands characteristic of a high  $\beta$ -sheet content together with new bands which could be specific for the carbohydrate and its association with the protein. Our results suggest that ROA spectra of intact glycoproteins may contain information about both protein and carbohydrate conformation and the mutual influence on each other's stability and conformation.

**Keywords:** Raman optical activity; glycoprotein conformation; vibrational optical activity

Raman optical activity (ROA), which refers to the measurement of vibrational optical activity by means of a small difference in the intensity of Raman scattering from chiral molecules in right and left circularly polarized incident light<sup>1</sup>, is proving to be an incisive new probe of biopolymer conformations in aqueous solution<sup>2-4</sup>. Since both carbohydrates and proteins are excellent samples for ROA studies, it is natural to enquire whether ROA can provide new information on glycoproteins, especially in view of the fact that intact glycoproteins are hard to study using conventional physical methods such as X-ray crystallography, two-dimensional nuclear magnetic resonance and electronic circular dichroism (CD). In this paper, we present and briefly discuss the first ROA spectrum of a glycoprotein.

We chose to study orosomucoid ( $\alpha_1$ -acid glycoprotein) because it is readily available and has a high carbohydrate content (~40%). Orosomucoid has a molecular weight of ~41 000 and is built from a single chain of 181 amino acid residues with five hetero-oligosaccharides attached by *N*-glycosidic bonds to asparagine residues<sup>5</sup>. Our sample, from human blood serum, was supplied by Sigma (Cohn fraction VI, 99%), and was run on the ROA instrument described previously<sup>6,7</sup>. Figure 1 shows the Raman and ROA spectra obtained at a concentration of 40 mg ml<sup>-1</sup> (1 mM) in 200 mM acetate buffer at pH 5.4.

Although a little noisy, it is immediately apparent that the ROA spectrum does indeed contain useful information. First, the sharpness of several of the protein ROA bands suggests that the biopolymer is unusually rigid. Second, we can see from the broad positive band at ~1060 cm<sup>-1</sup>, the strong sharp positive band at ~1313 cm<sup>-1</sup> and the negative-positive couplet with a

crossover at ~1660 cm<sup>-1</sup> that the protein has a high  $\beta$ -sheet content<sup>8</sup>; this contrasts with a recent study of orosomucoid which suggested that there might be only a small excess of  $\beta$ -sheet over  $\alpha$ -helix structures<sup>8</sup>. There is also a strong sharp negative ROA band at ~1245 cm<sup>-1</sup>, which also appears in insulin, ribonuclease A and lysozyme, where we have suggested that it might originate in a particular type of 'loop' with

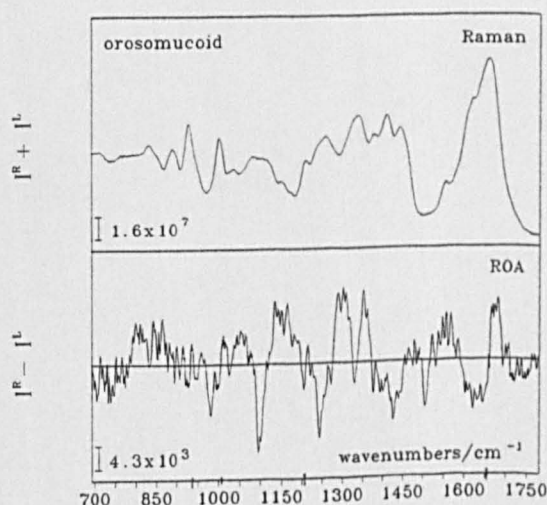


Figure 1 The backscattered Raman ( $I^R + I^L$ ) and ROA ( $I^R - I^L$ ) spectra of human blood serum orosomucoid in 200 mM acetate buffer (pH 5.4). Experimental conditions: laser wavelength = 514.5 nm; laser power at the sample = 600 mW; spectral band width = 10 cm<sup>-1</sup>; recording time = 20 h

\* To whom correspondence should be addressed

local order<sup>3</sup>. However, in addition, there are two striking ROA features which do not appear in any of the previous proteins studied to date and which could be specific for the carbohydrate and its association with the protein.

The first of these new features is a large couplet, negative and sharp at lower wavenumber and positive at high wavenumber, centred at  $\sim 1120\text{ cm}^{-1}$ . Early infra-red studies<sup>9</sup> of a glycopeptide of orosomucoid identified contributions from the carbohydrate in the range  $\sim 1000\text{--}1200\text{ cm}^{-1}$ . Our previous disaccharide ROA studies<sup>10</sup> indicate that this couplet might originate in C–O–C stretch vibrations of the glycosidic link. Furthermore, the polysaccharide laminarin displays large ROA signals in this range<sup>7</sup> which might be associated with its helical secondary structure.

The second new feature is a large sharp positive ROA band at  $\sim 1360\text{ cm}^{-1}$ . We have previously identified negative ROA bands in this region as originating in  $\beta$ -turn vibrations<sup>4</sup>, and, since it is thought that carbohydrates are often attached at  $\beta$ -turns<sup>11</sup>, we might speculate that this positive ROA band is associated with  $\beta$ -turns at the points of attachment.

Although insufficient data are available from this first preliminary glycoprotein ROA study to make definitive conclusions, it nonetheless appears that the ROA spectra of intact glycoproteins contain information about both the protein and the carbohydrate conformations and the mutual influence on each other's conformation and stability. An important aspect of future studies will be the use of linkage-specific endoglycosidase enzymes to cleave the glycans from the polypeptide so that a comparison can be made of the ROA spectrum of the complete glycoprotein with those of the separate carbohydrate and protein parts.

Recently, Urbanova *et al.* used the complementary technique of vibrational circular dichroism (VCD) to study a glycoprotein, glucoamylase<sup>12</sup>. Their results

provide another example of the potential value of measurement of vibrational optical activity in this area. They found that the conventional electronic CD analysis of the protein secondary structure was quantitatively in error due to interference by sugar residues. The VCD analysis was thought to be more reliable because it avoids such interference; however, unlike ROA, it provides no information on the carbohydrate conformation.

### Acknowledgements

The authors are grateful to the Science and Engineering Research Council and the Wellcome Trust for research grants, and to the Deutsche Forschungsgemeinschaft for a Research Fellowship for L. H. (Habilitationstipendium II C1-He 1588/3-1 & 3-2).

### References

- 1 Barron, L.D. 'Molecular Light Scattering and Optical Activity'. Cambridge University Press, Cambridge, 1982
- 2 Barron, L.D. and Hecht, L. in 'Advances in Spectroscopy, Vol. 21: Biomolecular Spectroscopy Part B' (Eds R.J.H. Clark and R.E. Hester) Wiley, Chichester, 1993, p. 235
- 3 Wen, Z.Q., Hecht, L. and Barron, L.D. *J. Am. Chem. Soc.* 1994, 116, 443
- 4 Wen, Z.Q., Hecht, L. and Barron, L.D. *Protein Sci.* 1994, 3, 435
- 5 Schmid, K., Kaufmann, H., Isemura, S., Bauer, F., Emura, J., Motoyama, T. *et al. Biochemistry* 1973, 12, 2711
- 6 Hecht, L., Barron, L.D., Gargaro, A.R., Wen, Z.Q. and Hug, W. *J. Raman Spectrosc.* 1992, 23, 401
- 7 Bell, A. F., Hecht, L. and Barron, L.D. *J. Raman Spectrosc.* 1993, 24, 633
- 8 Karpenko, V., Sinkorova, L. and Kodicek, M. *Coll. Czech. Chem. Commun.* 1992, 57, 641
- 9 Marshall, W.E. and Porath, J. *J. Biol. Chem.* 1965, 240, 209
- 10 Bell, A.F., Hecht, L. and Barron, L.D. *J. Am. Chem. Soc.* 1994, 116, 5155
- 11 Bentima, J.J. *Bioscience Rep.* 1986, 6, 709
- 12 Urbanova, M., Pancoska, P. and Keiderling, T.A. *Biochim. Biophys. Acta* 1993, 1203, 290





## Vibrational Raman optical activity of Biopolymers

L. D. Barron, L. Hecht, S. J. Ford, A. F. Bell and G. Wilson

Chemistry Department, The University, Glasgow G12 8QQ, United Kingdom

Vibrational Raman optical activity measurements can now provide a wealth of new information about the structure and conformation of biopolymers in aqueous solution. Typical results are exemplified here for proteins by bovine serum albumin in H<sub>2</sub>O and D<sub>2</sub>O solution, and for polysaccharides by laminarin.

### 1. INTRODUCTION

Raman optical activity (ROA) measures a small difference in the intensity of vibrational Raman scattering from chiral molecules in right and left circularly polarized incident light.<sup>1,2</sup> Recent advances in instrumentation based on backscattering, holographic notch filters and CCD detection have enhanced the sensitivity to the level necessary to provide ROA spectra of biological molecules in aqueous solution. ROA is an incisive new probe of biopolymer solution conformation:<sup>3–6</sup> this is because only those few local vibrational coordinates within a complicated normal mode which sample the skeletal chirality most directly make significant contributions to the associated ROA intensity, thereby generating characteristic band patterns which are usually much simpler than the parent Raman band patterns and which carry information about absolute configuration and conformation.

### 2. PEPTIDES AND PROTEINS

Vibrations of the peptide backbone are associated with four main regions of the Raman spectrum: the backbone C $\alpha$ –C stretch region at  $\sim 870$ – $950$  cm<sup>–1</sup>, the backbone C $\alpha$ –N stretch region at  $\sim 1020$ – $1150$  cm<sup>–1</sup>, the extended amide III region at  $\sim 1230$ – $1350$  cm<sup>–1</sup>, and the amide I region at  $\sim 1645$ – $1680$  cm<sup>–1</sup>. Signals usually appear in all four regions in most protein ROA spectra,<sup>3,4</sup> which show similarities with the ROA spectra of alanyl peptide oligomers.<sup>5</sup> In addition to clear signatures from extended secondary structures such as  $\alpha$ -helix and  $\beta$ -sheet, there appear to be prominent signatures from loops and turns<sup>3,4</sup> which have great potential value in the study of tertiary structure and dynamics of proteins.

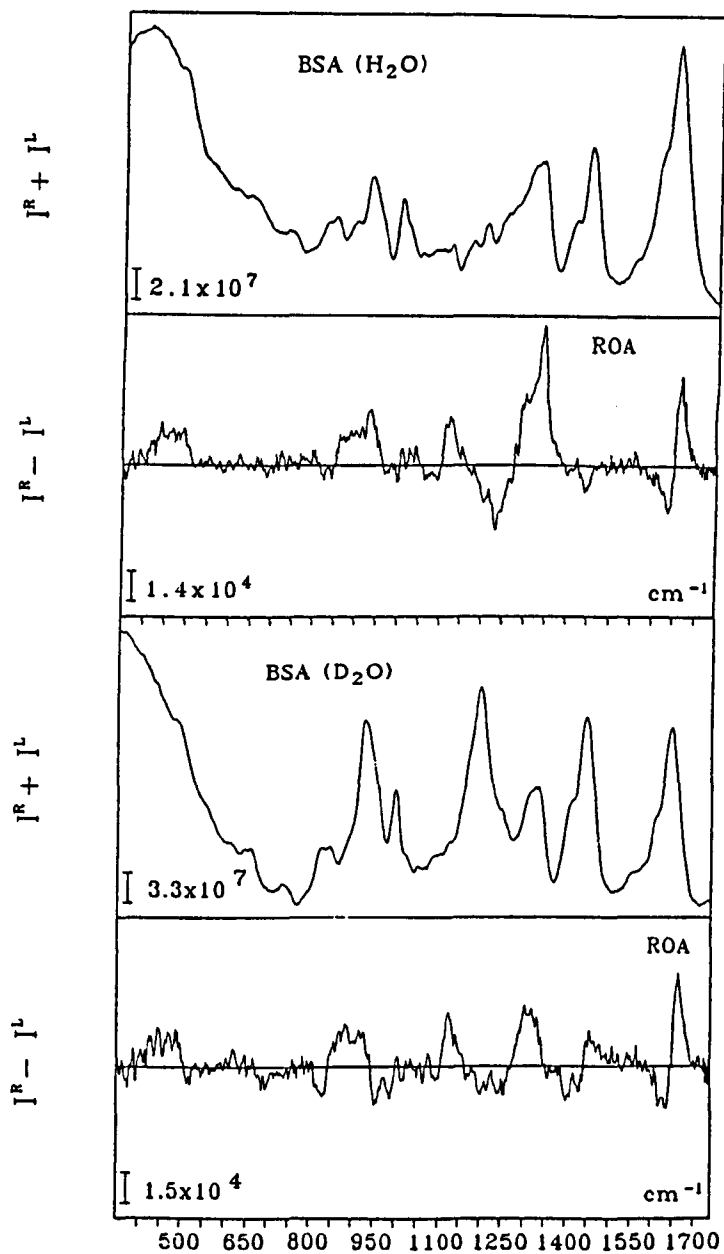


Figure 1. Backscattered Raman ( $I^R + I^L$ ) and ROA ( $I^R - I^L$ ) spectra of bovine serum albumin in hydrated and deuterated buffer (pH = 5.4).

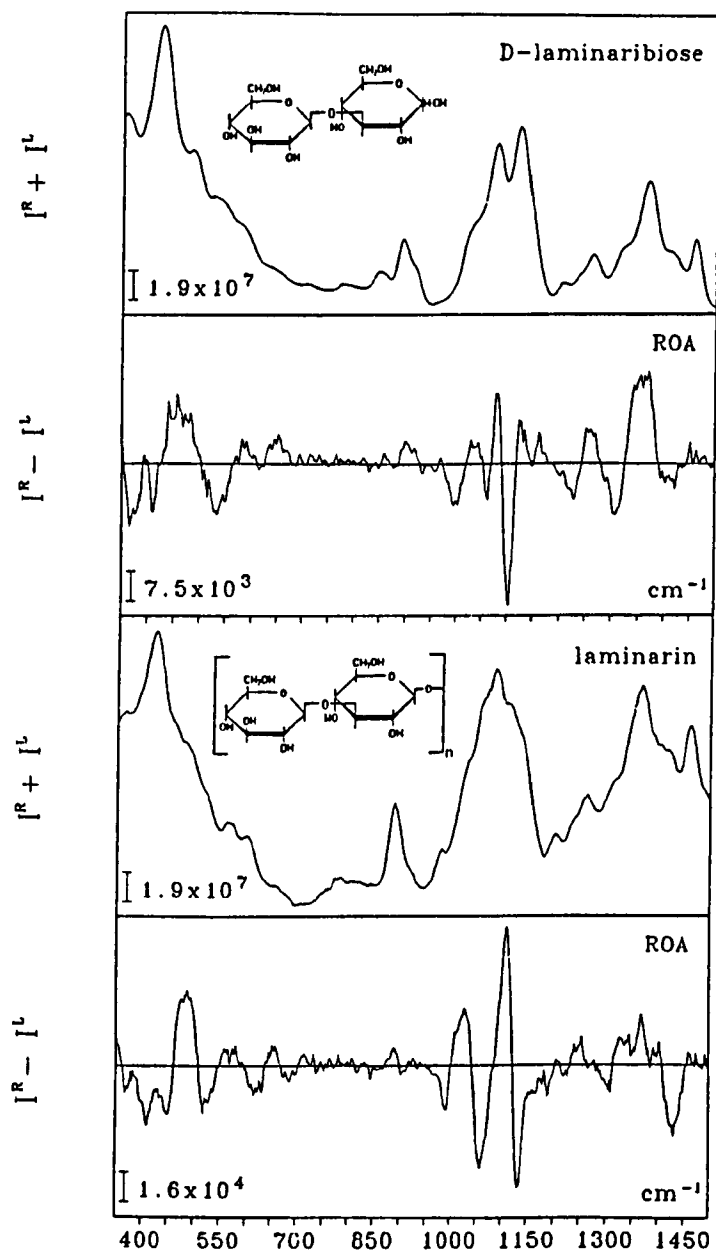


Figure 2. Backscattered Raman and ROA spectra of D-laminaribiose in  $\text{H}_2\text{O}$ , and of laminarin in tris buffer ( $\text{pH} = 7.5$ ).

Comparison of ROA spectra in  $\text{H}_2\text{O}$  and  $\text{D}_2\text{O}$  solution are very useful. The extended amide III region is particularly informative because replacement of N-H by N-D removes the crucial N-H deformations leaving just  $\text{C}_\alpha\text{-H}$  deformations to dominate the normal modes here so that the ROA is now generated mostly by the local chiral environment of the  $\alpha$ -carbon.<sup>5</sup> Bovine serum albumin (BSA) provides a good example (Figure 1). BSA is highly  $\alpha$ -helical with the remaining structure made up of loops, many of which are double and rigid on account of the large number of adjacent disulphide bridges. Most of the NH protons in the  $\alpha$ -helix structures would not have exchanged when the ROA was measured. The most striking change is the disappearance of the strong positive ROA band at  $\sim 1340\text{ cm}^{-1}$  and of the negative ROA band at  $\sim 1245\text{ cm}^{-1}$  which confirms their assignment to loop structures.<sup>3</sup> The backbone ROA features outside of the amide III region in  $\text{H}_2\text{O}$  which persist in  $\text{D}_2\text{O}$  solution can be assigned to  $\alpha$ -helix vibrations: in particular the positively-biased amide I couplet and the positive backbone skeletal stretch bands at  $\sim 1125$  and  $900\text{ cm}^{-1}$ . The positive ROA band at  $\sim 400\text{ cm}^{-1}$  might originate in  $\alpha$ -helix carbonyl deformations.

### 3. CARBOHYDRATES AND POLYSACCHARIDES

ROA spectra of carbohydrates in aqueous solution give information about the  $\alpha$ - or  $\beta$ -anomer, the pattern of OH substituents, the conformation of exocyclic  $\text{CH}_2\text{OH}$  groups, and the nature of the glycosidic link.<sup>6</sup> There also appears to be information about extended secondary structure in polysaccharides.

Figure 2 shows the ROA spectrum of laminarin, a  $\beta(1-3)$  linked polysaccharide of D-glucose, together with that of the corresponding  $\beta(1-3)$  disaccharide D-laminaribiose. The two ROA spectra are similar below  $\sim 1050\text{ cm}^{-1}$ . But in the region between  $\sim 1050$  and  $1150\text{ cm}^{-1}$ , where CO stretches involving the glycosidic oxygen contribute to the normal modes, the signs of the ROA signals are reversed on going from the dimer to the polymer which we believe is due to the formation of triple helical structures. The large differences above  $1200\text{ cm}^{-1}$  might reflect the pattern of hydrogen bonding between  $\text{CH}_2\text{OH}$  groups along the helix.

### REFERENCES

1. L. D. Barron, *Molecular Light Scattering and Optical Activity*, Cambridge University Press, Cambridge, 1982.
2. L. D. Barron and L. Hecht, in *Circular Dichroism: Principles and Applications* (K. Nakanishi, N. Berova and R. W. Woody, eds.), VCH Publishers, New York, 1994, p. 179.
3. Z. Q. Wen, L. Hecht and L. D. Barron, *J. Am. Chem. Soc.*, 116 (1994) 443.
4. Z. Q. Wen, L. Hecht and L. D. Barron, *Protein Science*, 3 (1994) 435.
5. S. J. Ford, Z. Q. Wen, L. Hecht and L. D. Barron, *Biopolymers*, 34 (1994) 303.
6. A. F. Bell, L. Hecht and L. D. Barron, *J. Am. Chem. Soc.*, 116 (1994) 5155.

## Vibrational Raman Optical Activity of $\alpha$ -Lactalbumin: Comparison with Lysozyme, and Evidence for Native Tertiary Folds in Molten Globule States

Gary Wilson, Steven J. Ford, Alan Cooper, Lutz Hecht, Zai Q. Wen and Laurence D. Barron\*

Chemistry Department  
The University, Glasgow  
G12 8QQ, UK

Proteins in aqueous solution are now accessible to Raman optical activity (ROA) measurements, which provide an incisive new probe of secondary and tertiary structure illustrated here by a study of bovine  $\alpha$ -lactalbumin. The room-temperature ROA spectrum of native bovine  $\alpha$ -lactalbumin is similar to that of native hen egg-white lysozyme except for features attributable to differences in the loop regions: in particular, a positive ROA band at  $\sim 1338\text{ cm}^{-1}$  assigned to conformationally homogeneous loop structure, possibly with local order corresponding to  $3_{10}$ -helix, has more than double the intensity in  $\alpha$ -lactalbumin compared with lysozyme. This is consistent with the two proteins having similar secondary structure but different local details in the tertiary fold. ROA measurements on  $\alpha$ -lactalbumin at pH 2.0 over a range of temperatures have provided a new perspective on the molten globule state. Thus at  $35^\circ\text{C}$  ROA reveals the presence of some secondary structure but an almost complete loss of the tertiary loop structure; whereas at  $2^\circ\text{C}$  the ROA spectrum is almost identical with that of the native protein, which is strong evidence that virtually all of the secondary structure and the tertiary backbone fold persist, albeit within a looser framework associated with increased solvent exposure and change of environment of many of the side-chains as evidenced by an increase in noise and bandwidth of some of the ROA signals together with aromatic fluorescence and near-UV circular dichroism signals characteristic of the molten globule state. Our sample of acid  $\alpha$ -lactalbumin at  $2^\circ\text{C}$  therefore appears to be an archetypal example of Ptitsyn's "native-like" molten globule, having a fixed native-like tertiary fold but with loss of tight packing of the side-chains; whereas at  $35^\circ\text{C}$  it is a "disordered" molten globule. At  $20^\circ\text{C}$  the acid molten globule appears to retain highly native-like secondary structure but with most of the tertiary fold already lost. A calcium-free sample of  $\alpha$ -lactalbumin at neutral pH displayed a broad cooperative transition between native and molten globule states at  $\sim 15^\circ\text{C}$ , with the latter state showing similar but somewhat degraded tertiary loop ROA signatures to the native protein. In both the acid and apo molten globule states the ROA signatures of the secondary structure and the tertiary loops showed a gradual change with temperature.

© 1995 Academic Press Limited

\*Corresponding author

**Keywords:** Raman optical activity;  $\alpha$ -lactalbumin; lysozyme; calcium binding; molten globule state

Abbreviations used: VCD, vibrational circular dichroism; ROA, Raman optical activity; NAG, N-acetylglucosamine; FTIR, Fourier transform infrared spectroscopy.

### Introduction

Vibrational optical activity has greatly enhanced the potential value of vibrational spectroscopy in

the study of chiral molecules in general and biopolymers in particular, since the observables have a unique sensitivity to molecular conformation (Nafie *et al.*, 1994). The vibrational circular dichroism (VCD) approach, which measures a small difference in the absorbance of left and right circularly polarized infrared radiation by chiral molecules, is now well established as a valuable probe of protein secondary structure (Keiderling & Pancoska, 1993; Pancoska *et al.*, 1994). Until recently, lack of instrument sensitivity has restricted application of the complementary technique of vibrational Raman optical activity (ROA), which measures a small difference in the intensity of Raman scattering from chiral molecules in right and left circularly polarized incident visible light (Barron, 1982; Diem, 1993; Barron & Hecht, 1994; Nafie & Che, 1994), to more favourable samples. However, thanks to new developments in instrumentation (Hecht *et al.*, 1992; Hecht & Barron, 1994; Hug, 1994), ROA can now be measured routinely on a large range of biological molecules in aqueous solution, and provides a new perspective on their solution structure (Wen, 1992; Barron & Hecht, 1993; Barron *et al.*, 1994, 1995). Proteins show particularly rich ROA spectra which, in addition to signatures characteristic of secondary backbone and side-group conformation, contain signatures that may be characteristic of loops and turns (Wen *et al.*, 1994a,b) so that the tertiary fold and its loss under various conditions can be studied. This, coupled with the fact that ROA measurements can be performed just as easily in  $\text{H}_2\text{O}$  and  $\text{D}_2\text{O}$  with no restriction on the size of the protein, makes ROA attractive for tackling many current problems in protein chemistry.

The protein  $\alpha$ -lactalbumin is present in the milk of most mammals where it combines with galactotransferase to switch its catalytic function from the production of *N*-acetylglucosamine to the production of lactose. Its structure and function, especially in relation to lysozyme, have attracted much interest and discussion (McKenzie & White, 1991; Sugai & Ikeguchi, 1994). In particular, although the X-ray crystal structures of lysozyme and  $\alpha$ -lactalbumin are very similar, other techniques appear to indicate significant differences in the aqueous solution structures. Furthermore,  $\alpha$ -lactalbumin has some interesting characteristics not possessed by most lysozymes, such as its ability to bind calcium and to support a molten globule state. Unfortunately, multidimensional NMR, which can usually provide more detailed information about the solution structure of smaller proteins than any other technique, has not yet produced a complete solution structure of native  $\alpha$ -lactalbumin on account of poor signal quality due to aggregation and dynamic fluctuations (Alexandrescu *et al.*, 1992). Other spectroscopic techniques have been applied intensively to the problem, but some of the results are ambivalent. ROA studies presented here shed new light on the relationship between the solution structures of lysozyme and  $\alpha$ -lactalbumin.

There is considerable current interest in stable non-native states of proteins that are intermediate between the native folded state and the denatured state (Dobson, 1994). The best known of these is the "molten globule" state adopted by  $\alpha$ -lactalbumin both at low pH and on removal of calcium at neutral pH, and by a number of other globular proteins when the native structure is disrupted in a relatively mild way, which is usually taken to have a compact globularity with native-like secondary structure but unfolded tertiary structure (Kuwajima, 1989). However, we show here that ROA measurements over a range of temperatures provides compelling evidence for the existence of some native-like tertiary folds, which it has been suggested might sometimes be present in molten globules in addition to the secondary structure elements (Ptitsyn, 1992).

In a previous paper, we presented an ROA study of hen egg-white lysozyme under various conditions, which provided new insights into the details of protein structure and dynamics in aqueous solution contained in an ROA spectrum (Ford *et al.*, 1995). This work provides a springboard for the ROA study of  $\alpha$ -lactalbumin presented here.

## Results and Discussion

### Structure of $\alpha$ -lactalbumin

We first briefly review the main features of the structure of  $\alpha$ -lactalbumin in order to facilitate discussion of its ROA spectra. The bovine form of  $\alpha$ -lactalbumin studied here has a molecular mass of 14.2 kDa and contains 123 amino acid residues. The structures of baboon and human  $\alpha$ -lactalbumin have both been determined from X-ray crystallography (Acharya *et al.*, 1989, 1991) and are very similar to that of lysozyme. Unfortunately the X-ray structure of bovine  $\alpha$ -lactalbumin itself has not been determined, but NMR studies (Alexandrescu *et al.*, 1992) indicate that its solution structure is similar to that of the human protein, which in turn has a crystal structure similar to that of baboon  $\alpha$ -lactalbumin (Acharya *et al.*, 1991). The structure of baboon  $\alpha$ -lactalbumin is shown schematically in Figure 1. There are four  $\alpha$ -helices, which make up one domain encompassing the amino and carboxy-terminal segments, and a small amount of double-stranded  $\beta$ -sheet that, together with a long loop, makes up much of a second domain. There are, in addition, four stretches of  $3_{10}$ -helix, three in the  $\alpha$ -helix domain and one in the second domain. The structure is stabilized by four disulphide bridges (6-120, 28-111, 61-77 and 73-91). The locations of the secondary structure elements, as taken from McKenzie & White (1991), are given in Table 1. The disulphide bridges between residues 61 and 77, and 73 and 91 generate a double loop that is of special interest. The overall length of this region and of the two components of the double loop are the same as in the equivalent region of lysozyme; however, the second loop in  $\alpha$ -lactalbumin encompasses the  $\text{Ca}^{2+}$  binding site, whereas hen egg-white lysozyme has

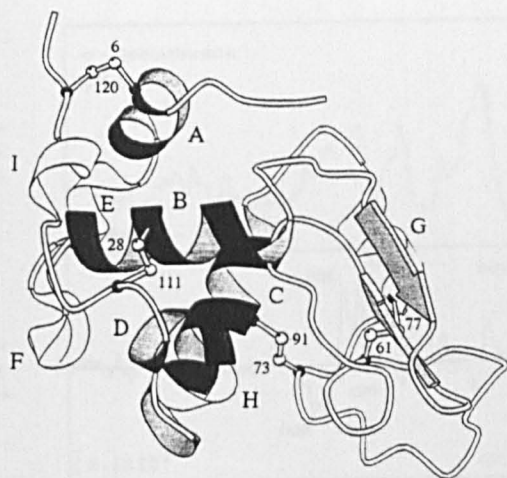


Figure 1. A schematic representation of the crystal structure of baboon  $\alpha$ -lactalbumin (from Acharya *et al.*, 1989). The Figure was drawn using the program MOLSCRIPT (Kraulis, 1991).

little affinity for binding metal ions. NMR studies (Alexandrescu *et al.*, 1992) have shown that the environments of the aromatic residues in bovine and human  $\alpha$ -lactalbumin are similar, and that the solution and crystal structures of the human protein are alike as evidenced by the similarity in the environments of the aromatic residues in the two states. Also, an early study that compared the conventional Raman spectrum of bovine  $\alpha$ -lactalbumin in the crystal and in solution concluded that the backbone conformations were very similar (Yu, 1974).

#### General features of the $\alpha$ -lactalbumin ROA spectrum: comparison with lysozyme

The backscattered Raman and ROA spectra of bovine  $\alpha$ -lactalbumin in acetate buffer at pH 5.4 are shown in the top third of Figure 2. The middle third shows those of hen egg-white lysozyme in the same buffer for comparison. Vibrations of the peptide backbone in proteins are usually associated with four main regions of the Raman spectrum (Tu, 1986): the C<sup>α</sup>-C stretch region  $\sim$ 870 to 950  $\text{cm}^{-1}$ , the C<sup>α</sup>-N stretch region  $\sim$ 1020 to 1150  $\text{cm}^{-1}$ , the amide III region  $\sim$ 1230 to 1310  $\text{cm}^{-1}$ , which is supposed to involve mainly the in-phase combination of largely the N-H in-plane deformation with the C<sup>α</sup>-N stretch, and the amide I region  $\sim$ 1630 to 1700  $\text{cm}^{-1}$  which arises predominantly from the C=O stretch. In fact, Diem *et al.* (1992) have shown that, in small

peptides, the amide III region involves much more mixing between the N-H and C<sup>α</sup>-H deformations than previously supposed, and should be extended to at least 1340  $\text{cm}^{-1}$ , which is supported by a detailed ROA study of a series of alanyl peptide oligomers (Ford *et al.*, 1994). This extended amide III region is particularly important for protein ROA work because the coupling between N-H and C<sup>α</sup>-H deformations is very sensitive to the geometry and generates a rich and informative ROA band structure. Side-group vibrations, especially those involving tryptophan, tyrosine, phenylalanine and CH<sub>2</sub> and CH<sub>3</sub> groups, also generate characteristic Raman bands (Tu, 1986) some of which show ROA signals. Bands in the amide II region  $\sim$ 1510 to 1570  $\text{cm}^{-1}$  (Diem, 1993), which originate in the out-of-phase combination of largely the N-H in-plane deformation with a small amount of C<sup>α</sup>-N stretch, are not usually observed in the conventional Raman spectra of proteins, but are prominent in the infrared and give rise to useful VCD spectra (Gupta & Keiderling, 1992); however, as discussed below, the amide II might be important in protein ROA studies.

The backbone skeletal stretch region has been shown to contain ROA bands originating in  $\alpha$ -helix (Wen *et al.*, 1994a) and  $\beta$ -sheet (Wen *et al.*, 1994b). The ROA spectrum of  $\alpha$ -lactalbumin is very similar to that of lysozyme in this region, showing positive intensity in the range  $\sim$ 880 to 960  $\text{cm}^{-1}$ , which can be assigned to  $\alpha$ -helix modes (and perhaps also the positive ROA band at  $\sim$ 855  $\text{cm}^{-1}$ , which is not shown by lysozyme), and a positive band at  $\sim$ 1006  $\text{cm}^{-1}$ , which might originate in  $\beta$ -sheet modes although contributions from aromatic side-groups cannot be ruled out. A second  $\alpha$ -helix ROA band structure appears in both proteins as a couplet with a small negative (often labile) component at  $\sim$ 1095  $\text{cm}^{-1}$  and a larger positive component at  $\sim$ 1120  $\text{cm}^{-1}$ . Small differences between the ROA spectra of the two proteins in this region might be attributed to differences in the details of the secondary structure such as different lengths of  $\alpha$ -helical segments and the larger amount of  $\beta$ -sheet in lysozyme. Despite these small differences the general similarity of the overall ROA fingerprints of the two proteins in the backbone skeletal stretch region suggests that they have similar secondary structure ( $\alpha$ -helix and  $\beta$ -sheet) content in aqueous solution.

Unlike VCD in the amide I region, which is very sensitive to different types of protein structure (Keiderling & Pancoska, 1993; Pancoska *et al.*, 1994), ROA in the amide I region is not particularly informative. Both  $\alpha$ -helix and  $\beta$ -sheet appear to give

Table 1. Secondary structure elements in baboon  $\alpha$ -lactalbumin (McKenzie & White, 1991)

Structure	Residues
$\alpha$ -Helix	5-11 (A), 23-34 (B), 86-99 (C), 105-109 (D)
3 <sub>10</sub> -Helix	12-16 (E), 17-21 (F), 76-82 (G), 101-104 (H), 115-119 (I)
$\beta$ -Sheet	40-43, 47-50



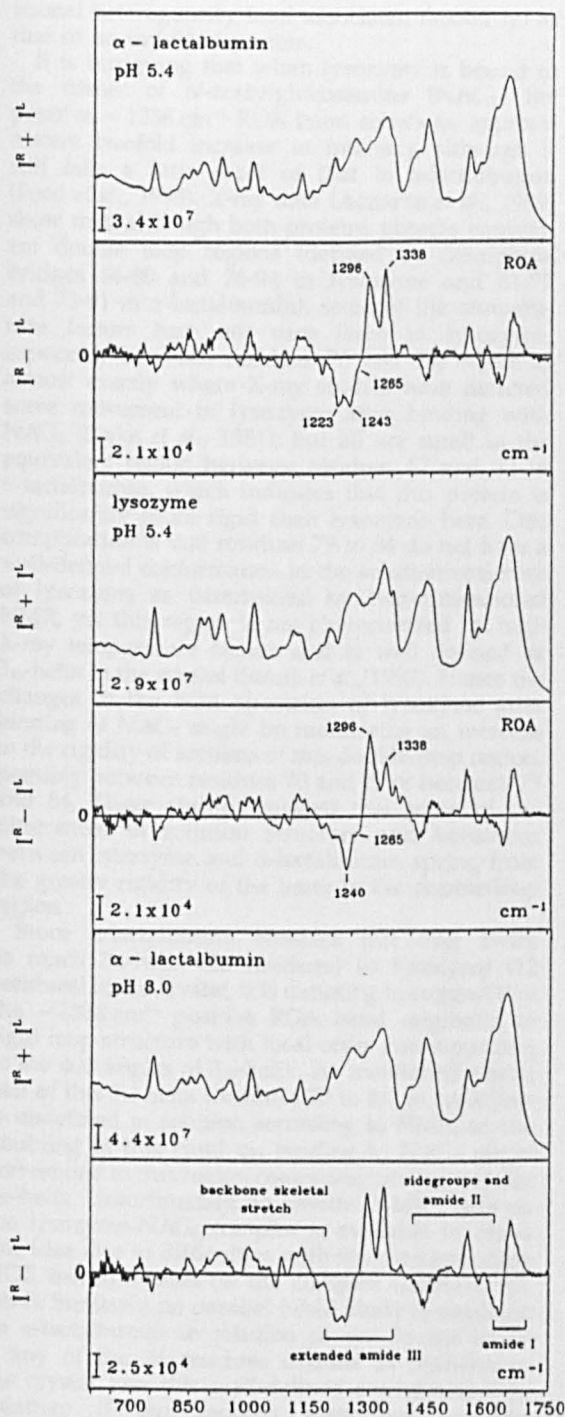


Figure 2. The backscattered Raman ( $I^R + I^L$ ) and ROA ( $I^R - I^L$ ) spectra of bovine  $\alpha$ -lactalbumin (top), and of hen egg-white lysozyme (middle), both in acetate buffer at pH 5.4; and of bovine  $\alpha$ -lactalbumin in Tris buffer at pH 8.0 (bottom). All were measured at room temperature ( $\sim 20^\circ\text{C}$ ). An approximate division of the ROA spectrum into distinct regions is indicated, together with positions of important ROA bands in the extended amide III region.

amide I couplets, negative at low wavenumber and positive at high: this couplet is usually conservative in  $\beta$ -sheet proteins, but has more intensity at higher wavenumber in  $\alpha$ -helix proteins. Apart from a small couplet overlapping the low-frequency end (*vide infra*), the appearance of the amide I ROA couplet in  $\alpha$ -lactalbumin is very similar to that in lysozyme, being consistent with a superposition of  $\alpha$ -helix and  $\beta$ -sheet couplets and also contributions from loops and turns (Ford *et al.*, 1995). So the amide I ROA is also consistent with a similar secondary structure content in the two proteins.

The most interesting region for protein ROA studies is the extended amide III. Although the overall appearance of the ROA spectrum here is similar in most proteins studied to date, namely a broad couplet negative at low wavenumber and positive at high, there is considerable structure that can be highly informative. Both  $\alpha$ -lactalbumin and lysozyme show a similar strong sharp positive peak at  $\sim 1296\text{ cm}^{-1}$  with a negative shoulder at  $\sim 1265\text{ cm}^{-1}$  assigned to  $\alpha$ -helix vibrations (Wen *et al.*, 1994a), which again supports a similar  $\alpha$ -helix content in the two proteins, but there is no evidence of the sharp positive band at  $\sim 1313\text{ cm}^{-1}$  shown by  $\beta$ -sheet proteins (Wen *et al.*, 1994b) presumably because the low  $\beta$ -sheet content of both generates only small ROA intensity here, which is swamped by the other intense bands in the vicinity. However, there are also significant differences between the ROA spectra of  $\alpha$ -lactalbumin and lysozyme in the extended amide III region that might shed light on the differences in solution structure between the two proteins. As suggested previously (Ford *et al.*, 1995), the sharp negative ROA band shown by lysozyme at  $\sim 1240\text{ cm}^{-1}$  (recorded at  $\sim 1245\text{ cm}^{-1}$  on the earlier instrument) and also a sharp positive ROA band at  $\sim 1338\text{ cm}^{-1}$  (recorded earlier at  $\sim 1340\text{ cm}^{-1}$ ), the latter appearing prominently in the ROA spectrum of bovine serum albumin, might originate in two distinct types of loop (or at least irregular) structure that are rigid and relatively homogeneous. The same  $\sim 1338\text{ cm}^{-1}$  positive ROA band appears in  $\alpha$ -lactalbumin but, strikingly, it is more than twice as intense than in lysozyme, which suggests that  $\alpha$ -lactalbumin contains two to three times as much of the corresponding type of rigid loop structure. Furthermore, the negative ROA band at  $\sim 1240\text{ cm}^{-1}$  in lysozyme is replaced in  $\alpha$ -lactalbumin by two bands, one peaking at  $\sim 1243\text{ cm}^{-1}$  and the other at  $\sim 1223\text{ cm}^{-1}$ , which suggests differences of local detail in other loop regions. In addition, our previous work on lysozyme (Ford *et al.*, 1995) reported that an unfolded form produced by reducing all the disulphide bonds showed a large unstructured couplet negative at  $\sim 1235\text{ cm}^{-1}$  and positive at  $\sim 1300\text{ cm}^{-1}$ , and we suggested that a similar couplet, which often underlies the amide III region with varying intensity in the ROA spectra of native proteins including those of  $\alpha$ -lactalbumin and lysozyme, might originate in loop and end-chain structure with the same type of conform-



ational heterogeneity (and associated flexibility) as that of an unfolded protein.

It is intriguing that when lysozyme is bound to the trimer of *N*-acetylglucosamine (NAG<sub>3</sub>) the positive  $\sim 1338\text{ cm}^{-1}$  ROA band shows an approximately twofold increase in intensity, although it still falls a little short of that in  $\alpha$ -lactalbumin (Ford *et al.*, 1995). X-ray data (Acharya *et al.*, 1989) show that, although both proteins possess equivalent double loop regions (defined by disulphide bridges 64-80 and 76-94 in lysozyme and 61-77 and 73-91 in  $\alpha$ -lactalbumin), some of the temperature factors here are very large in lysozyme, especially between residues 70 and 75, which is almost exactly where X-ray studies have detected some movement in lysozyme after binding with NAG<sub>3</sub> (Blake *et al.*, 1981); but all are small in the equivalent region between residues 67 and 72 in  $\alpha$ -lactalbumin, which indicates that this protein is significantly more rigid than lysozyme here. One complication is that residues 79 to 84 do not have a well-defined conformation in the solution structure of lysozyme as determined by two-dimensional NMR, yet this region is not characterized by high X-ray temperature factors and is well defined as  $3_{10}$ -helix in the crystal (Smith *et al.*, 1993). Hence the changes in the ROA spectrum of lysozyme after binding of NAG<sub>3</sub> might be monitoring an increase in the rigidity of sections of this double-loop region, possibly between residues 70 and 75 or between 79 and 84. These studies suggest that some of the differences in solution structure and behaviour between lysozyme and  $\alpha$ -lactalbumin spring from the greater rigidity of the latter in the double-loop region.

Since  $\alpha$ -lactalbumin contains just over twice as much  $3_{10}$ -helix (26 residues) as lysozyme (12 residues) in the crystal, it is tempting to suggest that the  $\sim 1338\text{ cm}^{-1}$  positive ROA band originates in rigid loop structure with local order corresponding to the  $\phi, \psi$  angles of  $3_{10}$ -helix. As mentioned above, half of this  $3_{10}$ -helix (residues 79 to 84) in lysozyme is undefined in solution according to NMR, so the doubling of this band on binding to NAG<sub>3</sub> might correspond to this region coalescing into the nascent  $3_{10}$ -helix. Unfortunately, no parallel NMR study on the lysozyme-NAG<sub>3</sub> complex is available to check this idea due to difficulties with the corresponding NOE measurements on the complex (Lumb *et al.*, 1994). Similarly, no parallel NMR study is available on  $\alpha$ -lactalbumin in solution so we do not know if any of the 26 residues defined as  $3_{10}$ -helix in the crystal lose this well-defined conformation in solution: if our mooted assignment of the  $\sim 1338\text{ cm}^{-1}$  ROA band is correct, then since only six residues in lysozyme survive as  $3_{10}$ -helix in solution, the observation of over twice, but certainly not four times, the intensity in this band in  $\alpha$ -lactalbumin could suggest that a number of the 26 residues defined as  $3_{10}$ -helix in the crystal might lose this conformation in solution. However, it is unlikely that these would be in the equivalent region to lysozyme, namely residues 76 to 82, since this forms

part of the calcium binding loop and so is expected to be especially rigid.

Side-group bands can also be identified in the  $\alpha$ -lactalbumin ROA spectrum. This protein contains four tryptophan residues that are responsible, among other things, for the strong positive ROA associated with the strong Raman band at  $\sim 1555\text{ cm}^{-1}$  and the weak negative ROA associated with the weaker Raman band at  $\sim 1580\text{ cm}^{-1}$ . The strong Raman band at  $\sim 1451\text{ cm}^{-1}$  with a shoulder at  $\sim 1420\text{ cm}^{-1}$  originates in side-group  $\text{CH}_2$  and  $\text{CH}_3$  deformations, with a contribution from tryptophan (mostly the NH bend of the indole ring) to the  $\sim 1420\text{ cm}^{-1}$  shoulder (Harada & Takeuchi, 1986), and is associated with an ROA couplet large and negative at low wavenumber with a small broad positive component extending beyond the high wavenumber limit of the parent Raman band. Although no significant conventional Raman band intensity is seen at wavenumbers associated with the small broad positive ROA component, it is known that the frequencies of the amide II peptide backbone vibration and of certain aromatic residue vibrations occur in this region (they become very strong in surface-enhanced Raman spectra: Hu *et al.*, 1995), so the observed ROA band structure arises from coupling of side-group  $\text{CH}_2$  and  $\text{CH}_3$  deformations with these backbone and side-group modes. Side-group ROA features similar to all of these are observed in lysozyme (Ford *et al.*, 1995); however, there are some small but significant differences. One feature present in the ROA spectrum of  $\alpha$ -lactalbumin but not of lysozyme is a small sharp couplet, positive at low wavenumber and negative at high, with the crossover point at  $\sim 1615\text{ cm}^{-1}$  near a shoulder on the low-wavenumber side of the amide I Raman band that is assigned to aromatic modes, including tryptophan (Lord & Yu, 1970; Harada & Takeuchi, 1986). Another difference is that the ROA structure associated with the  $\text{CH}_2$  and  $\text{CH}_3$  deformations Raman band is shifted by  $\sim 15\text{ cm}^{-1}$  to lower wavenumber in lysozyme. Hence the ROA data suggest that some differences exist in the conformations of a number of the aromatic and aliphatic side-groups in the two proteins.

#### Native-like and disordered acid molten globule states

Figure 3 shows the backscattered Raman and ROA spectra of  $\alpha$ -lactalbumin in glycine buffer at pH 2.0, under which conditions it is said to adopt the molten globule state (Kuwajima, 1989; Ptitsyn, 1992), measured at temperatures of 2°C, 20°C and 35°C.

Looking first at the ROA spectrum at 35°C shown at the bottom of Figure 3, we can recognize the hallmarks expected for a molten globule. Thus evidence that some (but not all) of the secondary structure persists is provided by the similarity of the signatures in the backbone skeletal stretch region to those in the spectrum of the native protein

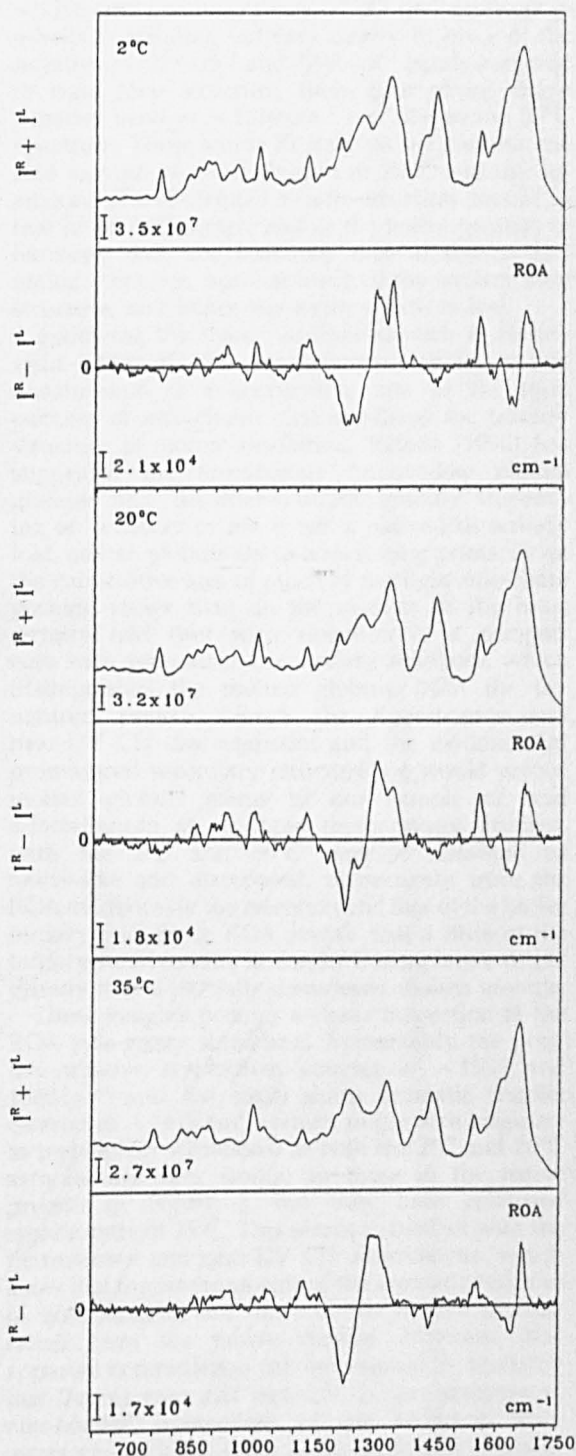


Figure 3. The backscattered Raman and ROA spectra of bovine  $\alpha$ -lactalbumin in glycine buffer at pH 2.0 measured at 2°C (top), 20°C (middle) and 35°C (bottom).

shown in Figure 2 but with reduced intensity and loss of certain detail; while the replacement of the ROA band structure in the extended amide III

region by a conservative couplet with a sharp negative peak at  $\sim 1230\text{ cm}^{-1}$  and a slightly broader positive peak at  $\sim 1300\text{ cm}^{-1}$ , very similar to that shown by unfolded (reduced) lysozyme (Ford *et al.*, 1995), suggests that most of the tertiary fold is lost with the loop structure now having a high degree of conformational heterogeneity and associated mobility. The amide I ROA couplet shows a narrowing and is shifted by  $\sim 8\text{ cm}^{-1}$  to higher wavenumber much like in unfolded lysozyme, which we can also associate with the increase in conformational heterogeneity (and loss of some secondary structure). The fluorescence ( $\lambda_{\text{exc}} = 295\text{ nm}$ ) of this sample, taken at 35°C, showed the expected red shift of the maximum from  $\sim 334\text{ nm}$  in the native protein at pH 8.0 to 346 nm with a small increase in intensity (e.g. Lala & Kaul, 1992); and the near-UV CD at 35°C showed the expected loss of the negative intensity seen in the native protein (Kuwajima, 1989). Such changes in fluorescence and near-UV CD are usually ascribed to changes in the environments and conformations of some of the aromatic residues. Thus the fluorescence red shift reflects a change of the tryptophan residues to a more polar environment associated with increased exposure to water, while the increase in intensity suggests the removal of tertiary interactions that quench the fluorescence in the native state (Hayer-Hartl *et al.*, 1994). The loss of the near-UV CD is usually ascribed to a time-averaged symmetrization of the environments of the aromatic groups due to slowly fluctuating tertiary structure (Dolgikh *et al.*, 1981).

Turning next to the 2°C ROA spectrum shown at the top of Figure 3, it is striking just how similar this is to that of the native protein in Figure 2. The secondary structure signatures in the backbone skeletal stretch and amide I regions, together with the  $\alpha$ -helix amide III couplet negative at  $\sim 1265$  and positive at  $\sim 1296\text{ cm}^{-1}$ , are so close in intensity and detail to those of the native protein that we are led to the conclusion that the secondary structure content is highly native-like. The presence of the full complement of amide III loop signatures (negative at  $\sim 1223$  and  $1243\text{ cm}^{-1}$  and positive at  $\sim 1338\text{ cm}^{-1}$ ) with full intensity further suggests that the native tertiary fold is present. Nonetheless, at 2°C this acid sample of  $\alpha$ -lactalbumin still exhibits the same red-shifted fluorescence and loss of near-UV CD as the 35°C sample, showing that the environments of the aromatic residues remain like those of a molten globule; and the broadening and increase in noise of the amide III loop signatures does suggest a loosening of the tertiary fold.

The ROA spectrum of the 20°C sample shown in the middle of Figure 3 reveals that the secondary structure content is still highly native-like (from the backbone skeletal stretch and amide I regions) but that interesting changes have taken place in the tertiary fold (from the extended amide III region). Thus, relative to the native protein (Figure 2), the positive  $\sim 1338\text{ cm}^{-1}$  band has lost most (but not all) of its intensity; whereas the couplet positive at

$\sim 1296\text{ cm}^{-1}$  and negative at  $\sim 1265\text{ cm}^{-1}$  assigned to  $\alpha$ -helix is standing out very clearly. In place of the negative  $\sim 1223\text{ cm}^{-1}$  and  $1243\text{ cm}^{-1}$  bands assigned to rigid loop structure, there is a strong sharp negative band at  $\sim 1230\text{ cm}^{-1}$  just like in the  $35^\circ\text{C}$  spectrum. These amide III features suggest that the acid sample of  $\alpha$ -lactalbumin at  $20^\circ\text{C}$  contains an amount of well-defined  $\alpha$ -helix structure similar to that of the  $2^\circ\text{C}$  sample and of the native protein, in harmony with the backbone skeletal stretch and amide I features, but that much of the tertiary loop structure, and hence the tertiary fold, is lost.

Following the theory of Shakhnovich & Finkelstein (1989a,b), that the primary step in protein denaturation is a cooperative loss of the tight packing of side-chains that stabilizes the tertiary structure at atomic resolution, Ptitsyn (1992) has suggested the terminology "native-like molten globule" and "disordered molten globule" depending on whether or not it has a native-like tertiary fold, molten globule status now resting primarily on the cooperative loss of much of the tight side-chain packing rather than on the absence of the basic tertiary fold (but with retention of a compact core with pronounced secondary structure, which distinguishes the molten globule from the denatured protein). From the fluorescence and near-UV CD characteristics and the evidence for pronounced secondary structure we would accord molten globule status to our sample of acid  $\alpha$ -lactalbumin at all three temperatures studied, with the  $2^\circ\text{C}$  and  $35^\circ\text{C}$  samples classified as native-like and disordered, respectively, from the ROA evidence for the retention and loss of the native tertiary fold. Since ROA reveals that a little of the tertiary fold remains in the  $20^\circ\text{C}$  sample we might classify it as a partially disordered molten globule.

These insights prompt a closer inspection of the ROA side-group signatures. Remarkably, the positive-negative tryptophan couplet at  $\sim 1555$  and  $1580\text{ cm}^{-1}$  and the small sharp aromatic couplet centred at  $\sim 1615\text{ cm}^{-1}$  (which might also originate in tryptophan vibrations) in both the  $2^\circ\text{C}$  and  $20^\circ\text{C}$  samples are very similar to those in the native protein in Figure 2, but they have collapsed significantly at  $35^\circ\text{C}$ . This seems to conflict with the fluorescence and near-UV CD observations, which show that the environments of the aromatic residues in both samples are those of the molten globule rather than the native protein. However, this apparent contradiction can be resolved by realizing that fluorescence and near-UV CD are sensitive to non-bonded interactions of the aromatic chromophores with the other residues, nearby parts of the peptide backbone and solvent molecules making up their environments, whereas ROA probes the local geometry of the aromatic residues relative to the peptide backbone via coupling of vibrational coordinates of the residues with those of the connecting covalent bonds. The ROA data therefore suggest that the tryptophan residues in the  $2^\circ\text{C}$  and  $20^\circ\text{C}$  acid molten globules retain a local rigidity similar to these residues in the native protein, even

though the global environments have changed. This conclusion accords with fluorescence polarization measurements of rotational relaxation times, which show that, at  $20^\circ\text{C}$ , the tryptophan residues in acid molten globule  $\alpha$ -lactalbumin are, if anything, even more restricted than in the native state (Dolgikh *et al.*, 1981).

Although the aromatic side-group ROA signatures remain well defined in the  $2^\circ\text{C}$  and  $20^\circ\text{C}$  acid molten globule samples, there appear to be significant changes relative to the native protein in the aliphatic side-group ROA signatures in the range  $\sim 1400$  to  $1520\text{ cm}^{-1}$ , which have collapsed into a small negative feature on the low-wavenumber side with no hint of the weak positive higher-wavenumber components. This suggests that the coupling between the  $\text{CH}_2$  and  $\text{CH}_3$  deformations of the side-groups with vibrational coordinates of the backbone amide II mode and of aromatic residue modes are suppressed, which implies increased conformational mobility of the aliphatic side-groups. This accords with the idea that "there is sufficient space inside the molten globule state for the free movements of small and symmetric aliphatic groups, but not enough for the movement of larger and less symmetric aromatic groups" (Ptitsyn, 1992). It could be significant that the coupling of the amide II to vibrational coordinates of side-groups is very sensitive to hydrogen bonding (Krimm & Bandekar, 1986) in view of the recent suggestion that the cooperative breaking of the hydrogen bond network is the rate-limiting step in protein unfolding (Kiefhaber *et al.*, 1995).

Recently, Wu *et al.* (1995) concluded from a study of two variants of  $\alpha$ -lactalbumin that are molten globules in the absence of calcium, even at neutral pH, that the molten globule properties (more specifically, native-like molten globule properties) of  $\alpha$ -lactalbumin are confined to its  $\alpha$ -helical domain, while the  $\beta$ -sheet domain (which contains the calcium binding loop) is largely unstructured. Molten globule status was inferred from UV CD measurements, while the existence or otherwise of a native-like backbone topology was inferred indirectly from the probability of formation of specific disulphide pairings. However, this conclusion might be over-simplified, because our direct observations of ROA loop signatures suggests that the presence or otherwise of elements of the tertiary fold in the molten globule state can be very sensitive to temperature (for given conditions of pH and salt concentration). Thus while the ROA spectrum of the acid molten globule sample at  $20^\circ\text{C}$  could be consistent with a bipartite structure, that of the same sample at  $2^\circ\text{C}$  strongly suggests that the native-like backbone topology encompasses the entire polypeptide chain.

#### Calcium-free states

As well as under acid conditions,  $\alpha$ -lactalbumin is also said to adopt a molten globule state on removal

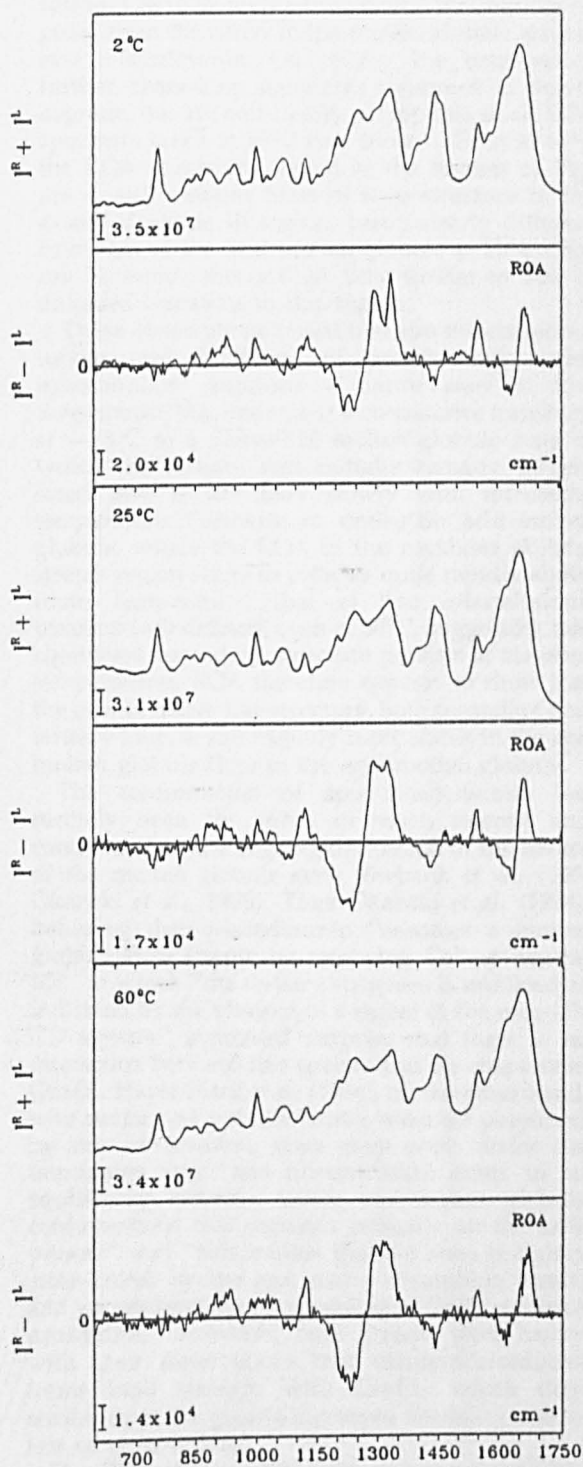


Figure 4. The backscattered Raman and ROA spectra of calcium-free (apo)  $\alpha$ -lactalbumin in Tris buffer at pH 8.0 measured at 2°C (top), 25°C (middle) and 60°C (bottom).

of calcium under conditions of low salt concentration at neutral pH (Kuwajima, 1989). We therefore studied calcium-free (apo)  $\alpha$ -lactalbumin in Tris

buffer at pH 8.0 over a range of temperatures to see what new insights ROA might provide into its structure. The backscattered Raman and ROA spectra taken at 2°C, 25°C and 60°C are shown in Figure 4. For comparison with the calcium-bound (holo) native form the ROA spectrum of the native protein was measured again in Tris buffer at pH 8.0 and is shown at the bottom of Figure 2.

The appearance of the 2°C apo ROA spectrum at the top of Figure 4 is very similar to that of the room temperature holo ROA spectrum (Figure 2), which strongly suggests that the overall structure of the apo form at 2°C is virtually unchanged from that of the holo form. In particular, all the loop signatures (the negative ROA bands at  $\sim 1223$  and  $1243\text{ cm}^{-1}$  and the strong positive  $\sim 1338\text{ cm}^{-1}$  ROA band) persist with similar intensity: since the calcium-binding loop is known to be one of the loops (that between residues 74 to 90) within the rigid double-loop region discussed above (Acharya *et al.*, 1989), our results indicate that the complete tertiary fold of holo  $\alpha$ -lactalbumin, including the calcium binding loop, persists at this temperature even when no calcium (or any other metal ion) is present. However, unlike the acid sample at 2°C, our sample of apo  $\alpha$ -lactalbumin is not a molten globule at 2°C as evidenced by the same  $\sim 334\text{ nm}$  fluorescence peak and negative near UV CD intensity shown by the native holo protein.

On raising the temperature, the fluorescence band intensity of apo  $\alpha$ -lactalbumin gradually increased and shifted towards the red, becoming stabilized at the maximum of  $\sim 346\text{ nm}$  expected for the molten globule state from previous studies (e.g. Permyakov *et al.*, 1981; Kronman *et al.*, 1981; Murukami *et al.*, 1982; Ewbank & Creighton, 1993) just below 25°C and above with a concomitant fall in the near UV CD intensity. On adding  $\text{CaCl}_2$  to the red-shifted sample, the fluorescence maximum and associated intensity reverted to that of the calcium-bound protein. More detailed studies of the temperature-dependence of the fluorescence and UV CD spectra of apo  $\alpha$ -lactalbumin, together with differential scanning calorimetry measurements, are in hand and will be reported elsewhere; but it appears that our sample of apo  $\alpha$ -lactalbumin undergoes a broad cooperative two-state transition centred at  $\sim 15^\circ\text{C}$ , being essentially 100% native below  $\sim 5^\circ\text{C}$  and 100% molten globule above  $\sim 25^\circ\text{C}$ .

The ROA spectrum at 25°C shown in the middle of Figure 4 still displays the loop signatures in the extended amide III region characteristic of the native tertiary fold, but all are significantly broader with more noise fluctuations than in the 2°C spectrum. It is interesting that the intensity of the positive  $\sim 1338\text{ cm}^{-1}$  ROA band has fallen a little relative to that in the 2°C sample (and also relative to the native holo protein): the loss of well-defined  $3_{10}$ -helix in residues 79 to 84 in lysozyme on going from the crystal to aqueous solution (*vide supra*) prompts the tentative notion that this intensity drop is monitoring a loss of  $3_{10}$ -helix in the equivalent



residues 76 to 82 within the calcium binding site on going from the native to the molten globule state of apo  $\alpha$ -lactalbumin. On raising the temperature further, these loop signatures continued to slowly degrade, but are still clearly perceptible in an ROA spectrum taken at 35°C (not shown). Even at 60°C the ROA spectrum, shown at the bottom of Figure 4, still contains hints of loop structure in the extended amide III region, being clearly different from that of the acid molten globule at 35°C (Figure 3), which shows ROA very similar to that of unfolded lysozyme in this region.

These observations reveal that apo  $\alpha$ -lactalbumin (under our conditions of protein and buffer concentration) supports a native state at low temperature that undergoes a cooperative transition at ~15°C to a native-like molten globule state in which the tertiary fold initially remains largely intact and is lost only slowly with increasing temperature. Furthermore, unlike the acid molten globule, where the ROA in the backbone skeletal stretch region starts to collapse quite rapidly above room temperature, that of apo  $\alpha$ -lactalbumin remains well defined, even at 60°C, suggesting that significant secondary structure persists at elevated temperatures. ROA therefore appears to show that the overall native-like structure, both secondary and tertiary loop, is significantly more stable in the apo molten globule than in the acid molten globule.

The conformation of apo  $\alpha$ -lactalbumin has recently been the focus of much interest and controversy in the ongoing discussion of the nature of the molten globule state (Ewbank *et al.*, 1995; Okazaki *et al.*, 1995). Thus Okazaki *et al.* (1994), believing that  $\alpha$ -lactalbumin "assumes a molten globule state simply by removing  $\text{Ca}^{2+}$  at neutral pH" in which "the tertiary structure is unfolded as indicated by the absence of a signal in the near-UV CD spectra" expressed surprise that there is no interaction between this species and the chaperonin GroEL. Hayer-Hartl *et al.* (1994), on the other hand, who performed a similar study, were not perplexed by this observation, since they were under the impression that "apo  $\alpha$ -lactalbumin exists in an equilibrium between native and molten globule conformations that depends critically on the salts present" and "this molten globule state is tightly cross-linked by the four native disulphide bonds, and we attribute the poor binding to GroEL to these constraints." However, both groups were happy with their observations that disulphide-reduced forms bind strongly with GroEL, which they attributed to the greatly increased flexibility due to loss of cross-linking.

Our ROA observations show that apo  $\alpha$ -lactalbumin is a native-like molten globule at ~25°C, the temperature at which the above GroEL binding studies were carried out, and it is the presence of the native-like tertiary fold, rather than cross-linking by the disulphide bonds, which probably explains why it does not bind to GroEL (recall that, according to the ROA observations presented in the previous section, the acid molten globule form of

$\alpha$ -lactalbumin shows a complete loss of the tertiary fold above ~25°C despite the disulphide bonds remaining intact). It is reassuring that Robinson *et al.* (1994), from studies of the hydrogen exchange kinetics of a reduced form of  $\alpha$ -lactalbumin bound to GroEL using electrospray ionization mass spectrometry, which showed that only disordered molten globules are recognized by GroEL, also concluded that apo  $\alpha$ -lactalbumin does not bind because it represents a more ordered molten globule state.

### Comparison with other spectroscopic studies

Although NMR has not been able to provide a complete solution structure of  $\alpha$ -lactalbumin, it has nonetheless been applied to certain aspects of the problem and it is gratifying that the conclusions from ROA and NMR are generally consistent. Thus the conclusion from NMR (Alexandrescu *et al.*, 1993; Chyan, *et al.*, 1993), that much of the native secondary structure, particularly that centred on the B and C helices, persists in the acid molten globule state of  $\alpha$ -lactalbumin is consistent with the ROA observations; and the failure of NMR to detect any significant residual native-like tertiary loop structure can be accommodated by the fact that the NMR was measured above 35°C where ROA shows an almost total loss of the native-like tertiary fold. NMR also showed no evidence in the acid molten globule for an ordered structure for either of the two major clusters of aromatic residues that in the native structure make up part of the hydrophobic core of the helical domain of the native protein, but a new smaller aromatic cluster was detected (Alexandrescu *et al.* 1993): this is consistent with the 35°C ROA spectrum (Figure 2) in which the aromatic side-group ROA bands, while still discernible, are weaker and more diffuse than in the native protein; however, since these aromatic ROA bands are strong and sharp in the 20°C spectrum where the secondary structure signatures look highly native-like, a native-like clustering might be observed if the NMR studies were repeated at lower temperature. Furthermore, from studies of urea-induced unfolding, where it was found that the unfolding transitions as monitored by changes in individual aromatic resonances do not coincide, NMR has shown that unfolding of the acid molten globule state of  $\alpha$ -lactalbumin is not a cooperative two-state process (Shimizu *et al.*, 1993), which accords with the ROA observations of gradual changes with temperature in the signatures of both secondary structure and tertiary loops within the acid molten globule state. Also, since the ROA spectra of apo  $\alpha$ -lactalbumin show clear evidence of native-like tertiary folds, even above room temperature, it is likely that loss of calcium does not lead to any significant change in the peptide backbone conformation (apart from a possible loss of 3<sub>10</sub>-helix in the calcium binding loop in the molten globule state) and so agrees with the conclusion from NMR studies (Berliner *et al.*, 1987) that the two forms are

"not drastically different in their three dimensional conformations". A recent VCD study has reached a similar conclusion (Keiderling *et al.*, 1994).

Using conventional FTIR measurements of amide I band structure in  $^2\text{H}_2\text{O}$  solution, Prestrelskii *et al.* (1991) concluded that removal of  $\text{Ca}^{2+}$  from holo  $\alpha$ -lactalbumin results in a local unfolding of the  $\text{Ca}^{2+}$ -binding loop. Despite reservations about possible ambiguities in the interpretation arising from the different hydrogen-deuterium exchange rates producing only partial exchange of backbone amide groups, and from the recent suggestion (Krimm & Reisdorf, 1994) that the suppression of collective modes on introducing irregularities into helices makes it dangerous to interpret component amide I peaks in terms of individual "standard" secondary structure contributions, this conclusion could be consistent with the ROA studies, which indicate a loss of  $3_{10}$ -helix possibly within the calcium binding loop at the "ambient temperature" (presumably above  $\sim 25^\circ\text{C}$ ) at which the FTIR measurements were carried out.

Studies based on a comparison of the amide I VCD band shapes of lysozyme and  $\alpha$ -lactalbumin in  $^2\text{H}_2\text{O}$  (Urbanova *et al.*, 1991) and in  $\text{H}_2\text{O}$  (Keiderling *et al.*, 1994) have suggested that the secondary structures of  $\alpha$ -lactalbumin and lysozyme in aqueous solution are much less similar than in the crystal. This contrasts with the conclusion from the present ROA study that  $\alpha$ -lactalbumin and lysozyme have quite similar secondary structures in aqueous solution, any differences originating mainly in local details of the tertiary fold with  $\alpha$ -lactalbumin having greater rigidity in the region of the calcium binding loop. The interpretation of these VCD results is the subject of ongoing investigations (Keiderling *et al.*, 1994).

Conventional electronic CD studies have had mixed fortunes in the study of  $\alpha$ -lactalbumin. For example, whereas electronic CD has not cast much light on whether or not significant differences exist between the solution structures of  $\alpha$ -lactalbumin and lysozyme (McKenzie & White, 1991), it has proved valuable in studies of the molten globule state since, as mentioned above, the far-UV CD shows the presence of significant secondary structure whereas the absence of near-UV CD shows that the environments of the aromatic residues are symmetrized (Kuwaitjima, 1989; Ptitsyn, 1992; Sugai & Ikeguchi, 1994; Fink, 1995). However, the ROA observations on acid and apo molten globule states of  $\alpha$ -lactalbumin described above suggest that the commonly held belief that the absence of a signal in the near-UV CD indicates that the tertiary structure is unfolded (Okazaki *et al.*, 1994) can be misleading in some situations. It is safer to accept that the presence of near-UV CD from aromatic residues "provides a specific and sensitive fingerprint of their native tertiary packing" (Christensen & Pain, 1994) without going on to conclude that loss of near-UV CD automatically implies a concomitant loss of the tertiary fold. The detailed mechanisms by which near-UV CD is

induced in aromatic chromophores in protein side-chains are still not fully understood (Woody, 1994) so it remains unclear exactly how the loss of the near-UV CD comes about in the molten globule state, especially in a highly ordered molten globule such as acid  $\alpha$ -lactalbumin at low temperature where ROA has shown the tertiary fold to be highly native-like and the tryptophan residues rigid.

## Concluding Remarks

On account of its ability to monitor both secondary structure, and tertiary loop structure, and to distinguish between conformationally homogeneous (rigid) and heterogeneous (flexible) elements of the latter, ROA has provided a fresh perspective on the solution structure of  $\alpha$ -lactalbumin and has helped to resolve some of the controversy generated by ambivalent results from other spectroscopic techniques.

Of particular significance is the ability of ROA together with complementary UV CD and fluorescence studies to distinguish between "native-like" molten globules and those with varying degrees of disorder in their tertiary fold and how this changes with temperature, together with its ability to recognize loss of rigidity in aliphatic and aromatic (tryptophan) side-group conformations separately. In addition to acid and apo  $\alpha$ -lactalbumin discussed here and the variants studied by Wu *et al.* (1995), two other examples of molten globules containing native-like tertiary backbone structure have been recognized recently. Thus Redfield *et al.* (1994) found from heteronuclear NMR studies on interleukin-4 that "despite the CD and fluorescence changes, the structure at low pH is highly native-like, and that a substantial global loss of tertiary structure has not occurred". In another NMR study, Feng *et al.* (1994) found that apocytochrome  $b_{562}$  has a topology similar to the holoprotein but with extensive solvation of the side-groups in the large binding pocket.

Our results therefore reinforce the emerging consensus view that non-native states of globular proteins encompass a broad range of conformations and degrees of folding and compactness having no single, unique conformation but rather a whole plethora of structures that range from being very similar to the native state to being substantially expanded and significantly unfolded (Fink, 1995). A similar view has been expressed by Dobson (1994). Thus, although removal of  $\text{Ca}^{2+}$  destabilizes the native structure of  $\alpha$ -lactalbumin, it appears to remain essentially intact at low temperature and neutral pH. Increase in temperature leads to a broad cooperative transition involving change in aromatic residue (tryptophan) environments, whilst maintaining much of the backbone secondary structure and tertiary fold. Reduction in pH, which prevents  $\text{Ca}^{2+}$  binding by protonation of the binding site ligands to give a net positively charged polypeptide, leads to a different sampling of conformational sub-space. At low temperature, in

acid, conformations congregate in regions resembling the native tertiary fold but with tryptophan residues still exposed to solvent water. With increase in temperature the polypeptide progressively explores ever greater regions of conformational space with no global cooperative changes involved.

The ROA evidence that molten globule states of  $\alpha$ -lactalbumin under certain conditions appear to support a virtually intact native fold adds weight to the view that they are important intermediates in protein folding pathways because the remaining search for the fully folded state need be over only a dramatically reduced area of conformational space (Dobson, 1995).

We are aware of possible problems associated with the rather high protein concentrations used that are necessary to achieve a good ROA signal-to-noise ratio with the current instrumentation. This might cause concern that aggregation, to which partially folded states of proteins are prone, is partially responsible for the stabilization of the tertiary loop structure seen in apo and low pH  $\alpha$ -lactalbumin under various conditions. Until further developments in instrumentation make it possible to measure protein ROA spectra at much lower concentrations, we can only point out that the observed ROA loop signatures from the same sample can range from fully defined to completely collapsed over certain temperature ranges, and that in the usual temperature realm of NMR studies ( $\sim 35^\circ\text{C}$ ) the deductions from ROA are fully consistent with those from NMR.

There remains the question of the precise nature of the rigid loop structures responsible for the ROA bands seen in the extended amide III regions and that have featured so prominently in this study. The strong possibility that the positive  $\sim 1338\text{ cm}^{-1}$  band originates in  $3_{10}$ -helix, which is consistent with all of our ROA observations to date on other polypeptides and proteins, and its utility in monitoring local  $3_{10}$ -helix unfolding, is highly topical in view of the potential importance of  $3_{10}$ -helix recently recognized in protein packing (Liu & Day, 1994) and in the folding pathway of helical peptides (Millhauser, 1995). However, the origin of the negative band at  $\sim 1240\text{ cm}^{-1}$  in lysozyme and of the negative bands at  $\sim 1223$  and  $\sim 1243\text{ cm}^{-1}$  in  $\alpha$ -lactalbumin remain obscure (except that they certainly do not originate in  $\alpha$ -helix or  $\beta$ -sheet). For the moment we shall leave these assignments open and hope that studies of other proteins together with results from *ab initio* computations of the ROA spectra of model polypeptide fragments over a range of conformations (Polavarapu & Deng, 1994) will provide further insight into the detailed local geometries of these and other loops revealed by ROA.

## Materials and Methods

### Materials

Protein solutions for ROA studies were prepared using the following buffers: 0.1 M sodium acetate, pH 5.4 (for

native lysozyme and  $\alpha$ -lactalbumin comparison); 0.1 M Tris, pH 8.0 (for calcium-free and native  $\alpha$ -lactalbumin comparison); and 0.1 M glycine, pH 2.0 (for  $\alpha$ -lactalbumin in the acid molten globule state) using Analar grade reagents and distilled, deionized water throughout. Protein concentrations were determined by measuring UV absorbance of diluted aliquots assuming an extinction coefficient (for  $\alpha$ -lactalbumin)  $E_{280\text{nm}} = 2.01\text{ ml mg}^{-1}\text{ cm}^{-1}$  (Gill & von Hippel, 1989). Protein solutions (typically  $\sim 70\text{ mg/ml}$ ) were prepared in small glass sample tubes, mixed with a little activated charcoal (pharmaceutical grade) to remove traces of fluorescing impurities, and centrifuged. Solutions were subsequently filtered through Millipore GV4 ( $0.22\text{ }\mu\text{m}$ ) filters directly into quartz microfluorescence cells, which were then again centrifuged gently (to clear any remaining dust particles) prior to mounting in the ROA instrument. Residual visible fluorescence from traces of impurities, which can give large background levels in Raman spectra, was allowed to "burn down" by leaving the sample to equilibrate in the laser beam for several hours before acquiring ROA data. This procedure, which is standard in the Raman spectroscopy of proteins (Carey, 1982), did not appear to harm our samples because their Raman spectra remained unchanged (denaturation is readily detected in Raman spectra). Our confidence in the integrity of the proteins following exposure to the laser beam during the ROA measurements was reinforced by finding that the lysozyme sample retained its full enzymatic activity (as measured by the standard Sigma procedure based on cell wall hydrolysis of *Micrococcus lysodeikticus*), and that the fluorescence characteristics of both the calcium-bound and calcium-free samples of  $\alpha$ -lactalbumin were unchanged.

Hen egg-white lysozyme and type I bovine  $\alpha$ -lactalbumin (reported to contain 1 to 2 mol Ca/mol protein) were obtained from Sigma Chemical Co. and used without further purification. Type III bovine  $\alpha$ -lactalbumin (Sigma), which is partially calcium-depleted, was further treated as follows to remove residual  $\text{Ca}^{2+}$  and other metal cations. Approximately 100 mg of type III  $\alpha$ -lactalbumin was dissolved in 5 ml of 20 mM EDTA/ammonia solution (pH 8.2), and stirred for about one hour at room temperature. This was then dialysed against  $5 \times 200\text{ ml}$  volumes of EDTA/ammonia, followed by extensive dialysis against deionized water ( $7 \times 3\text{ l}$ ) before freeze drying.

Tryptophan fluorescence measurements on diluted protein samples ( $0.05$  to  $0.1\text{ mg/ml}$ ) were performed over a  $0$  to  $35^\circ\text{C}$  temperature range in  $1\text{ cm}$  path-length quartz cuvettes,  $\lambda_{\text{exc}} = 295\text{ nm}$ , using a SPEX FluorMax spectrofluorimeter. The UV CD measurements were performed on a Jasco J-600 spectropolarimeter using cells of path-length  $0.02\text{ cm}$  (far-UV, concentrations  $\sim 0.6\text{ mg/ml}$ ) and  $0.05\text{ cm}$  or  $0.5\text{ cm}$  (near-UV, concentrations  $\sim 6\text{ mg/ml}$  or  $0.6\text{ mg/ml}$ ).

### ROA spectroscopy

The instrument used for the ROA measurements was an upgraded version of one described previously (Hecht *et al.*, 1992). It utilizes backscattering, which is essential for aqueous solution samples of biopolymers, and employs a single-grating spectrograph fitted with a backthinned CCD camera as detector and a holographic notch filter to block the Rayleigh line. The most important change has been the use of a new stigmatic spectrograph based on a novel transmission diffraction grating, which has provided an approximately fivefold increase in the

speed of ROA measurements with improved resolution (Hecht & Barron, 1994). ROA is measured by synchronizing the spectral acquisition with an electro-optic modulator used to switch the polarization of the incident argon ion laser beam between right and left circular at a suitable rate. The ROA spectra are presented in the form of a circular intensity difference (in analogue-to-digital converter counts)  $I^R - I^L$ , where  $I^R$  and  $I^L$  are the Raman-scattered intensities in right and left circularly polarized incident light. The conventional Raman intensities are presented as a corresponding circular intensity sum  $I^R + I^L$ . The experimental conditions were as follows: laser wavelength 514.5 nm; laser power at the sample  $\sim 700$  mW; spectral bandwidth  $10\text{ cm}^{-1}$ ; recording time about ten hours.

For ROA measurements at other than room temperature ( $\sim 20^\circ\text{C}$ ), air was blown over the sample cell using an FTS systems model TC-84 Air Jet Crystal Cooler, which can supply dry air over the range  $-85$  to  $+100^\circ\text{C}$ .

Small discrepancies (up to  $\sim 5\text{ cm}^{-1}$ ) in the quoted wavenumbers for some of the same ROA bands between this and earlier protein ROA papers originate in different calibration protocols used with different spectrographs and also in uncertainties in some of the ROA peak positions due to noise fluctuations.

## Acknowledgements

We thank the EPSRC and the BBSRC for research grants, the EPSRC for a Senior Fellowship for L.D.B. and a Studentship for G.W., and the Wellcome Trust for a Prize Studentship for S.J.F. We are grateful to Dr S. E. Radford for helpful discussions, to a referee for pointing out the significance of our ROA observations for current discussions of the nature of the molten globule state, and to Professor N. C. Price and Dr S. N. Kelley for performing the UV CD measurements on the BBSRC-funded facility at Stirling University.

## References

- Acharya, K. R., Stuart, D. I., Walker, N. P. C., Lewis, M. & Phillips, D. C. (1989). Refined structure of baboon  $\alpha$ -lactalbumin at  $1.7\text{ \AA}$  resolution. Comparison with c-type lysozyme. *J. Mol. Biol.* **208**, 99–127.
- Acharya, K. R., Ren, J., Stuart, D. I., Phillips, D. C. & Fenna, R. E. (1991). Crystal structure of human  $\alpha$ -lactalbumin at  $1.7\text{ \AA}$  resolution. *J. Mol. Biol.* **221**, 571–581.
- Alexandrescu, A. T., Broadhurst, R. W., Wormald, C., Chyan, C. L., Baum, J. & Dobson, C. M. (1992).  $^1\text{H}$ -NMR assignments and local environments of aromatic residues in bovine, human and guinea pig variants of  $\alpha$ -lactalbumin. *Eur. J. Biochem.* **210**, 699–709.
- Alexandrescu, A. T., Evans, P. A., Pitkeathly, M., Baum, J. & Dobson, C. M. (1993). Structure and dynamics of the acid-denatured molten globule state of  $\alpha$ -lactalbumin: a two-dimensional NMR study. *Biochemistry*, **32**, 1707–1718.
- Barron, L. D. (1982). *Molecular Light Scattering and Optical Activity*, Cambridge University Press, Cambridge.
- Barron, L. D. & Hecht, L. (1993). Biomolecular conformational studies with vibrational Raman optical activity. In *Advances in Spectroscopy*, (Clark R. J. H. & Hester, R. E., eds), vol. 21, pp. 235–266, Wiley, Chichester.
- Barron, L. D. & Hecht, L. (1994). Vibrational Raman optical activity: from fundamentals to biochemical applications. In *Circular Dichroism Principles and Applications* (Nakanishi, K., Berova, N. & Woody, R. W., eds), pp. 179–215, VCH Publishers, New York.
- Barron, L. D., Ford, S. J., Bell, A. F., Wilson, G., Hecht, L. & Cooper, A. (1994). Vibrational Raman optical activity of biopolymers. *Faraday Discuss.* **99**, 217–232.
- Barron, L. D., Hecht, L. & Bell, A. F. (1995). Vibrational Raman optical activity of biomolecules. In *Circular Dichroism: Conformational Analysis of Biomolecules* (Fasman, G. D., ed.), Plenum Press, New York, in the press.
- Berliner, L. J., Koga, K., Nishikawa, H. & Scheffler, J. E. (1987). High-resolution proton and laser photochemically induced dynamic nuclear polarization NMR studies of cation binding to bovine  $\alpha$ -lactalbumin. *Biochemistry*, **26**, 5769–5774.
- Blake, C. C. F., Cassels, R., Dobson, C. M., Poulsen, F. M., Williams, R. J. P. & Wilson, K. S. (1981). Structure and binding properties of hen lysozyme modified at tryptophan 62. *J. Mol. Biol.* **147**, 73–95.
- Carey, P. R. (1982). *Biochemical Applications of Raman and Resonance Raman Spectroscopies*, Academic Press, New York.
- Christensen, H. & Pain, R. H. (1994). The contribution of the molten globule model. In *Mechanisms of Protein Folding* (Pain, R. H., ed.), pp. 55–79, IRL Press, Oxford.
- Chyan, C.-L., Wormald, C., Dobson, C. M., Evans, P. A. & Baum, J. (1993). Structure and stability of the molten globule state of guinea-pig  $\alpha$ -lactalbumin: a hydrogen exchange study. *Biochemistry*, **32**, 5681–5691.
- Diem, M. (1993). *Modern Vibrational Spectroscopy*, Wiley, New York.
- Diem, M., Lee, O. & Roberts, G. M. (1992). Vibrational studies, normal coordinate analysis, and infrared VCD of alanylalanine in the amide III spectral region. *J. Phys. Chem.* **96**, 548–554.
- Dobson, C. M. (1994). Solid evidence for molten globules. *Curr. Biol.* **4**, 636–640.
- Dobson, C. M. (1995). Finding the right fold. *Nature struct. Biol.* **2**, 513–517.
- Dolgikh, D. A., Gilmanishin, R. I., Brazhnikov, E. V., Bychkova, V. E., Semisotnov, G. V., Venyaminov, S. Yu & Ptitsyn, O. B. (1981).  $\alpha$ -Lactalbumin: compact state with fluctuating tertiary structure? *FEBS Letters*, **136**, 311–315.
- Ewbank, J. J. & Creighton, T. E. (1993). Structural characterization of the disulphide folding intermediates of bovine  $\alpha$ -lactalbumin. *Biochemistry*, **32**, 3694–3707.
- Ewbank, J. J., Creighton, T. E., Hayer-Hartl, M. K. & Hartl, F. U. (1995). What is the molten globule? *Nature Struct. Biol.* **2**, 10.
- Feng, Y., Sligar, S. G. & Wand, A. J. (1994). Solution structure of apocytochrome  $b_{562}$ . *Nature Struct. Biol.* **1**, 30–35.
- Fink, A. L. (1995). Compact intermediate states in protein folding. *Annu. Rev. Biophys. Biomol. Struct.* **24**, 495–522.
- Ford, S. J., Wen, Z. Q., Hecht, L. & Barron, L. D. (1994). Vibrational Raman optical activity of alanyl peptide oligomers. *Biopolymers*, **34**, 303–313.
- Ford, S. J., Cooper, A., Hecht, L., Wilson, G. & Barron, L. D. (1995). Vibrational Raman optical activity of lysozyme: hydrogen-deuterium exchange, unfolding



- and ligand binding. *J. Chem. Soc. Faraday Trans.* 91, 2087–2093.
- Gill, S. C. & von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* 182, 319–326.
- Gupta, V. P. & Keiderling, T. A. (1992). Vibrational CD of the amide II band in some model polypeptides and proteins. *Biopolymers*, 32, 239–248.
- Harada, I. & Takeuchi, H. (1986). Raman and ultraviolet resonance Raman spectra of proteins and related compounds. In *Advances in Spectroscopy*, (Clark, R. J. H. & Hester, R. E., eds), vol. 13, pp. 113–175, Wiley, Chichester.
- Hayer-Hartl, M. K., Ewbank, J. J., Creighton, T. E. & Hartl, F. U. (1994). Conformational specificity of the chaperonin GroEL for the compact folding intermediates of  $\alpha$ -lactalbumin. *EMBO J.* 13, 3192–3202.
- Hecht, L. & Barron, L. D. (1994). Recent developments in Raman optical activity instrumentation. *Faraday Discuss.* 99, 35–47.
- Hecht, L., Barron, L. D., Gargaro, A. R., Wen, Z. Q. & Hug, W. (1992). Raman optical activity instrument for biochemical studies. *J. Raman Spectrosc.* 23, 401–411.
- Hu, J., Sheng, R. S., Xu, Z. S. & Zeng, Y. (1995). Surface enhanced Raman spectroscopy of lysozyme. *Spectrochim. Acta Ser. A*, 51, 1087–1096.
- Hug, W. (1994). Vibrational Raman optical activity comes of age. *Chimia*, 48, 386–390.
- Keiderling, T. A. & Pancoska, P. (1993). Structural studies of biological macromolecules using vibrational circular dichroism. In *Advances in Spectroscopy* (Clark, R. J. H. & Hester, R. E., eds), vol. 21, pp. 267–315, Wiley, Chichester.
- Keiderling, T. A., Wang, B., Urbanova, M., Pancoska, P. & Dukor, R. K. (1994). Empirical studies of protein secondary structure by vibrational circular dichroism and related techniques.  $\alpha$ -Lactalbumin and lysozyme as examples. *Faraday Discuss.* 99, 263–285.
- Kiefhaber, T., Labhardt, A. M. & Baldwin, R. L. (1995). Direct NMR evidence for an intermediate preceding the rate-limiting step in the unfolding of ribonuclease A. *Nature*, 375, 513–515.
- Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* 24, 946–950.
- Krimm, S. & Bandekar, J. (1986). Vibrational spectroscopy and conformation of peptides, polypeptides, and proteins. *Advan. Protein Chem.* 38, 181–364.
- Krimm, S. & Reisdorf, W. C., Jr (1994). Understanding normal modes of proteins. *Faraday Discuss.* 99, 181–197.
- Kronman, M. J., Sinha, S. K. & Brew, K. (1981). Characteristic of the binding of  $\text{Ca}^{2+}$  and other divalent metal ions to bovine  $\alpha$ -lactalbumin. *J. Biol. Chem.* 256, 8582–8587.
- Kuwajima, K. (1989). The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins: Struct. Funct. Genet.* 6, 87–103.
- Lala, A. K. & Kaul, P. (1992). Increased exposure of hydrophobic surface in molten globule state of  $\alpha$ -lactalbumin. *J. Biol. Chem.* 267, 19914–19918.
- Liu, D. J. & Day, L. A. (1994). Pfl virus structure: helical coat protein and DNA with paraxial phosphates. *Science*, 265, 671–674.
- Lord, R. C. & Yu, N.-T. (1970). Laser Raman spectroscopy of biomolecules. I. Native lysozyme and its constituent amino acids. *J. Mol. Biol.* 50, 509–524.
- Lumb, K. J., Cheetham, J. C. & Dobson, C. M. (1994).  $^1\text{H}$  Nuclear magnetic resonance studies of hen lysozyme-N-acetylglucosamine oligosaccharide complexes in solution. *J. Mol. Biol.* 235, 1072–1087.
- McKenzie, H. A. & White, F. H. (1991). Lysozyme and  $\alpha$ -lactalbumin: structure, function, and interrelationships. *Advan. Protein Chem.* 41, 173–316.
- Millhauser, G. L. (1995). Views of helical peptides—a proposal for the position of  $3_{10}$ -helix along the thermodynamic folding pathway. *Biochemistry*, 34, 3873–3877.
- Murakami, K., Andree, P. J. & Berliner, L. J. (1982). Metal ion binding to  $\alpha$ -lactalbumin species. *Biochemistry*, 21, 5488–5494.
- Nafie, L. A. & Che, D. (1994). Theory and measurement of Raman optical activity. *Advan. Chem. Phys.* 85(3) 105–149.
- Nafie, L. A., Yu, G.-S., Qu, X. & Freedman, T. B. (1994). Comparison of IR and Raman forms of vibrational optical activity. *Faraday Discuss.* 99, 13–34.
- Okazaki, A., Ikura, T., Nikaido, K. & Kuwajima, K. (1994). The chaperonin GroEL does not recognize apo- $\alpha$ -lactalbumin in the molten globule state. *Nature Struct. Biol.* 1, 439–446.
- Okazaki, A., Ikura, T. & Kuwajima, K. (1995). Reply to Ewbank *et al.*, 1995. *Nature Struct. Biol.* 2, 10–11.
- Pancoska, P., Bitto, E., Janota, V. & Keiderling, T. A. (1994). Quantitative analysis of vibrational circular dichroism spectra of proteins. *Faraday Discuss.* 99, 287–310.
- Permyakov, E. A., Yarmolenko, V. V., Kalinichenko, L. P., Morozova, L. A. & Burstein, E. A. (1981). Calcium binding to  $\alpha$ -lactalbumin: structural rearrangement and association constant evaluation by means of intrinsic protein fluorescence changes. *Biochem. Biophys. Res. Commun.* 100, 191–197.
- Polavarapu, P. L. & Deng, Z. (1994). Structural determinations using vibrational Raman optical activity: from a single peptide group to  $\beta$ -turns. *Faraday Discuss.* 99, 151–163.
- Prestrelski, S. J., Blyler, D. M. & Thompson, M. P. (1991). Effect of metal binding on the secondary structure of bovine  $\alpha$ -lactalbumin as examined by infrared spectroscopy. *Biochemistry*, 30, 8797–8804.
- Ptitsyn, O. B. (1992). In *Protein Folding* (Creighton, T. E., ed.), pp. 243–300, Freeman, New York.
- Redfield, C., Smith, R. A. G. & Dobson, C. M. (1994). Structural characterization of a highly-ordered 'molten globule' at low pH. *Nature Struct. Biol.* 1, 23–29.
- Robinson, C. V., Groß, M., Eyles, S. J., Ewbank, J. J., Mayhew, M., Hartl, F. U., Dobson, C. M. & Radford, S. E. (1994). Conformation of GroEL-bound  $\alpha$ -lactalbumin probed by mass spectrometry. *Nature*, 372, 646–651.
- Shakhnovich, E. I. & Finkelstein, A. V. (1989a). Theory of cooperative transitions in protein molecules. I. Why denaturation of globular protein is a first-order phase transition. *Biopolymers*, 28, 1667–1680.
- Shakhnovich, E. I. & Finkelstein, A. V. (1989b). Theory of cooperative transitions in protein molecules. II. Phase diagram for a protein molecule in solution. *Biopolymers*, 28, 1681–1694.
- Shimizu, A., Ikeguchi, M. & Sugai, S. (1993). Unfolding of the molten globule state of  $\alpha$ -lactalbumin studied by  $^1\text{H}$  NMR. *Biochemistry*, 32, 13198–13203.
- Smith, L. J., Sutcliffe, M. J., Redfield, C. & Dobson, C. M. (1993). Structure of hen lysozyme in solution. *J. Mol. Biol.* 29, 930–944.

- Sugai, S. & Ikeguchi, M. (1994). Conformational comparison between  $\alpha$ -lactalbumin and lysozyme. *Advan. Biophys.* 30, 37–84.
- Tu, A. T. (1986). Peptide backbone conformation and microenvironment of protein sidechains. In *Advances in Spectroscopy* (Clark, R. J. H. & Hester, R. E., eds), vol. 13, pp. 47–112. Wiley, Chichester.
- Urbanova, M., Dukor, R. K., Pancoska, P., Gupta, V. P. & Keiderling, T. A. (1991). Comparison of  $\alpha$ -lactalbumin and lysozyme using vibrational circular dichroism. Evidence for a difference in crystal and solution structures. *Biochemistry*, 30, 10479–10485.
- Wen, Z. Q. (1992). Raman optical activity of biological molecules, Doctoral thesis, Glasgow University.
- Wen, Z. Q., Hecht, L. & Barron, L. D. (1994a).  $\alpha$ -Helix and associated loop signatures in vibrational Raman optical activity spectra of proteins. *J. Am. Chem. Soc.* 116, 443–445.
- Wen, Z. Q., Hecht, L. & Barron, L. D. (1994b).  $\beta$ -sheet and associated turn signatures in vibrational Raman optical activity spectra of proteins. *Protein Sci.* 3, 435–439.
- Woody, R. W. (1994). Circular dichroism of peptides and proteins. In *Circular Dichroism Principles and Applications* (Nakanishi, K., Berova, N. & Woody, R. W., eds), pp. 473–496. VCH Publishers, New York.
- Wu, L. C., Peng, Z. & Kim, P. S. (1995). Bipartite structure of the  $\alpha$ -lactalbumin molten globule. *Nature Struct. Biol.* 2, 281–286.
- Yu, N.-T. (1974). Comparison of protein structure in crystals, in lyophilized state, and in solution by laser Raman scattering. *J. Am. Chem. Soc.* 96, 4664–4668.

Edited by A. R. Fersht

(Received 29 March 1995; accepted in revised form 26 September 1995)

# Vibrational Raman optical activity of $\alpha$ -helical and unordered poly(L-lysine)

Gary Wilson, Lutz Hecht and Laurence D. Barron\*

Chemistry Department, The University, Glasgow, UK G12 8QQ

Measurements of the vibrational Raman optical activity (ROA) spectra of poly(L-lysine) in  $H_2O$  and  $D_2O$  solutions, under conditions said to generate model  $\alpha$ -helical and unordered conformations, are reported. These provide new insights into the conformational elements actually present. In addition to several signatures characteristic of  $\alpha$ -helix, the  $\alpha$ -helical state (pH 11.0, 20°C) of a sample with  $\bar{M}_w = 26\,000$  shows a strong sharp positive ROA band at ca.  $1340\text{ cm}^{-1}$ , previously assigned to rigid loop structure with local order possibly that of  $3_{10}$ -helix, which suggests that the  $\alpha$ -helical sections are connected by segments of the corresponding loop. On increasing  $\bar{M}_w$  to 268 000 this loop signature grows to dominate the ROA spectrum with a concomitant decrease in the  $\alpha$ -helix bands. The unordered state (pH 3.0, 20°C) shows clear signatures of several distinct conformational elements, including  $\alpha$ -helix,  $\beta$ -strand and, possibly, left-handed helix. Under conditions which tend to disrupt secondary structure (5 mol  $l^{-1}$  NaCl solution heated to 50°C) the ROA spectrum of the unordered state shows a decrease in the intensity of the signature of the putative left-handed helix and an increase in bands assigned to  $\alpha$ -helix.

Determination of the solution structure of polypeptides and proteins remains a central problem in biophysical chemistry. The potential value of vibrational spectroscopy, both IR and Raman,<sup>1-3</sup> in this area has been enhanced by adding the new dimension of optical activity,<sup>4-6</sup> which confers considerably more sensitivity to molecular conformation. The vibrational circular dichroism (VCD) approach, which measures a small difference in absorbance of left and right circularly polarized IR radiation by chiral molecules, is now well established as a valuable probe of peptide and protein secondary structure.<sup>7</sup> Lack of sufficient instrument sensitivity has, until recently, restricted application of the complementary technique of vibrational ROA, which measures a small difference in the intensity of vibrational Raman scattering from chiral molecules in right and left circularly polarized incident light, to more favourable samples.<sup>8</sup> However, thanks to new developments in instrumentation,<sup>9-11</sup> ROA can now be measured routinely on a large range of biological molecules in aqueous solution and provides a fresh perspective on their solution structure.<sup>12-16</sup> Polypeptides and proteins show particularly rich ROA spectra which, in addition to signatures characteristic of secondary backbone and sidegroup conformation, also appear to contain signatures from loops and turns.<sup>17-20</sup> Preliminary ROA spectra of poly(L-lysine) in supposedly  $\alpha$ -helical and unordered conformations have been published previously and appear to contain new information about the details of the solution conformation.<sup>12,17</sup> In view of the importance of these states of poly(L-lysine) as models for some central features of polypeptide and protein conformation, and of the interest and controversy which still surrounds the precise nature of the unordered state and how it might differ from a true random coil,<sup>21</sup> we present here a more detailed study based on ROA data of higher quality obtained using more recent instrumentation.

The normal vibrational modes of polypeptides and proteins can be highly complex, with contributions from many skeletal and sidegroup local vibrational coordinates all of which can affect the intensity and frequency.<sup>2</sup> The main advantage of ROA relative to conventional Raman spectroscopy here is its ability to cut through some of the inherent complexity of the conventional vibrational spectrum of the biopolymer. This is because only those few local vibrational coordinates within a

complicated normal mode, which sample the skeletal chirality most directly, make significant contributions to the associated ROA intensity, thereby generating characteristic ROA band patterns which are usually much simpler and more informative than the parent Raman band patterns. Furthermore, unlike VCD which usually has to rely on  $D_2O$  solutions for clear signal identification, ROA spectra can be measured just as easily in  $H_2O$  and  $D_2O$  and a comparison of the two can be highly informative.<sup>16,19</sup>

Polypeptides which adopt  $\alpha$ -helix,  $\beta$ -sheet and unordered conformations under certain conditions of temperature and pH have long been used as models for spectroscopic identification of corresponding secondary structure elements in proteins: thus poly(L-lysine) supports an unordered state at low pH, an  $\alpha$ -helix state at high pH and low temperature, and a  $\beta$ -sheet state at high pH and high temperature.<sup>22,23</sup> Raman spectra have been reported for all three conformations in aqueous solution.<sup>24-26</sup> Unfortunately, experimental difficulties prevented us from obtaining the ROA spectrum of the  $\beta$ -sheet state, but high-quality ROA spectra of  $\alpha$ -helical and unordered poly(L-lysine) were measured which will be discussed together with the ROA spectrum of L-lysine.

## Experimental

Samples of poly(L-lysine) (hydrobromide,  $\bar{M}_w = 26\,000$  and 268 000 by viscosity) and L-lysine were obtained from Sigma, together with the  $D_2O$ . The L-lysine was dissolved in filtered deionised  $H_2O$  to a concentration of 2 mol  $l^{-1}$ . A solution of poly(L-lysine) with a concentration of 100 mg  $ml^{-1}$  was prepared in a 0.1 mol  $l^{-1}$  glycine buffer at room temperature at pH 3.0, under which conditions an unordered conformation is expected.<sup>22,23</sup> A corresponding solution in 5 mol  $l^{-1}$  NaCl was obtained by combining solutions of NaCl and the polypeptide in the same buffer at appropriate concentrations. Some difficulty was experienced in preparing a solution of  $\alpha$ -helical poly(L-lysine) at the required pH 11.0 at 100 mg  $ml^{-1}$ . The procedure finally adopted was to add dropwise to a weighed sample of poly(L-lysine) the appropriate number of moles, plus a 10% excess, of NaOH necessary to deprotonate all of the amide nitrogens at ca. 0°C in an ice bath. On warming the sample, it was found to form a gel (presumably

due to crosslinking of the  $\alpha$ -helical structure), so cold dry air was blown over the sample cell in order to create an environment of ca. 2°C during the recording of the ROA spectra. The D<sub>2</sub>O solutions were prepared in a similar fashion. Unfortunately the  $\beta$ -sheet form of poly(L-lysine) obtained by heating the high pH sample exists as a gel with large random birefringence which renders it unsuitable for ROA measurements.

The solutions were filtered through Millipore GV4 (0.22  $\mu$ m) filters into quartz microfluorescence cells, which were subsequently centrifuged and left to equilibrate for at least one day at the temperature of the measurement. Residual fluorescence was allowed to 'burn down' by leaving the sample in the laser beam for several hours before acquiring the ROA spectra.

The instrument used for the ROA measurements is a development of our original backscattering ROA instrument.<sup>9</sup> This was based on a single-grating astigmatic spectrograph fitted with a backthinned CCD detector with a holographic notch filter to block the Rayleigh line, ROA being measured using the incident circular polarization (ICP) strategy<sup>6</sup> by synchronizing the spectral acquisition with an electro-optic modulator used to switch the polarization of the incident argon ion laser beam between right and left circular at a suitable rate. This instrument has recently been upgraded<sup>10</sup> with the introduction of a stigmatic Kaiser Holospec VPT spectrograph which employs a novel transmission diffraction grating permitting light collection at much lower *F*-numbers which provides an approximately fivefold increase in speed compared with the spectrograph previously employed. The ROA spectra are presented in the form of a circular intensity difference (in analogue-to-digital converter counts)  $I^R - I^L$ , where  $I^R$  and  $I^L$  are the backscattered Raman intensities in right- and left-circularly polarized incident light. The conventional Raman spectra are presented as a corresponding circular intensity sum  $I^R + I^L$ . The experimental conditions were as follows: laser wavelength, 514.5 nm; laser power at the sample, ca. 600 mW; spectral resolution ca. 12 cm<sup>-1</sup>; recording time, ca. 3 h for the amino acid and 10 h for the polypeptides.

For ROA measurements at other than room temperature (ca. 20°C) air was blown over the sample cell using an FTS Systems Model TC-84 Air Jet Crystal Cooler which can supply dry air over the range -85 to +100°C.

## Results and Discussion

### L-Lysine

It is important to know whether intrinsic sidegroup ROA makes significant contributions to the polypeptide ROA spectra: the relatively long sidechains of the constituent amino acids could take up interesting chiral conformations with large associated ROA. So we first show and briefly discuss, with emphasis on the CH<sub>2</sub> modes, the Raman and ROA spectra (Fig. 1) of the amino acid L-lysine, since these have not been published previously.

A detailed experimental and *ab initio* theoretical study of L-alanine has been published,<sup>27,28</sup> followed by more refined *ab initio* studies,<sup>6,29</sup> together with an experimental study of some other simple amino acids.<sup>30</sup> There are many similarities between the ROA spectrum of L-lysine and those of L-alanine and some of the other simple amino acids in H<sub>2</sub>O in the range ca. 700–1200 cm<sup>-1</sup> which can be attributed to normal modes involving vibrational coordinates of the -CH(NH<sub>3</sub><sup>+</sup>)(CO<sub>2</sub>H) group. Most of the differences can be attributed to C–C stretch with CH<sub>2</sub> rock (at the lower end of this wavenumber range) and CH<sub>2</sub> twist (at the higher end) coordinates of the -(CH<sub>2</sub>)<sub>4</sub>- groups.<sup>31,32</sup>

The negative ROA band structure shown by L-lysine in H<sub>2</sub>O in the range ca. 1250–1340 cm<sup>-1</sup> is not shown by alanine, but prominent negative ROA bands appear in this range in L-serine where they were attributed to CH<sub>2</sub> twist and wag coordinates, together with COH bends.<sup>30</sup> It is likely that CH<sub>2</sub> twisting and wagging also contribute to these negative ROA bands in L-lysine, with significant contributions from NH deformations suggested by the changes observed in the ROA in this region in D<sub>2</sub>O solution.

The large positive band at ca. 1363 cm<sup>-1</sup> and the large negative band at ca. 1405 cm<sup>-1</sup> shown by L-lysine in H<sub>2</sub>O are similar to features in the ROA spectra of L-alanine, L-serine and L-threonine which have been attributed to C<sub>α</sub>-H deformations and to the symmetric CO<sub>2</sub><sup>-</sup> stretch, respectively, so the same assignments can be made for L-lysine. The positive ROA band at ca. 1445 cm<sup>-1</sup> can be assigned to the CH<sub>2</sub> scissors deformation.<sup>31,32</sup> The very similar appearance of the ROA spectrum of L-lysine in D<sub>2</sub>O in this region reinforces our assignments of these three ROA bands.

### General features of polypeptide vibrations

Vibrations of the backbone in polypeptides and proteins are usually associated with four main regions of the Raman spectrum:<sup>3,5,26,33</sup> the C<sub>α</sub>-C stretch region ca. 870–950 cm<sup>-1</sup>, the C<sub>α</sub>-N stretch region ca. 1020–1150 cm<sup>-1</sup>, the amide III region ca. 1230–1310 cm<sup>-1</sup> which is supposed to involve mainly the in-phase combination of largely the N–H in-plane deformation coupled with the C<sub>α</sub>-N stretch, and the amide I region ca. 1630–1700 cm<sup>-1</sup> which arises predominantly from the C=O stretch. In fact, Diem and co-workers<sup>3,34</sup> have shown that, in small peptides, the amide III region involves much more mixing between the N–H and C<sub>α</sub>-H deformations than previously supposed, and should be extended to at least 1340 cm<sup>-1</sup>. This is supported by a detailed ROA study of a series of alanyl peptide oligomers.<sup>35</sup> This extended amide III region is particularly important for polypeptide and protein ROA work because the coupling between N–H and C<sub>α</sub>-H deformations is very sensitive to the geometry and generates a rich and informative ROA band structure containing signatures from loops and turns in addition to secondary structure. Sidegroup vibrations, especially those involving tryptophan, tyrosine, phenylalanine and CH<sub>2</sub> and CH<sub>3</sub> groups, also generate characteristic Raman bands some of which show ROA signals. Bands in the amide II region ca. 1510–1570 cm<sup>-1</sup>, which originate in the out-of-phase combination of largely the N–H in-plane deformation with a small amount of C<sub>α</sub>-N stretch, are not usually observed in the conventional Raman spectra of peptides and proteins, but are prominent in the IR.<sup>5</sup> However, the amide II region becomes important in Raman and ROA studies in D<sub>2</sub>O solution.

### $\alpha$ -Helical poly(L-lysine)

The conventional Raman spectra of the  $\bar{M}_n = 26\,000$  sample of  $\alpha$ -helical poly(L-lysine) in H<sub>2</sub>O and D<sub>2</sub>O shown in Fig. 2 are virtually identical to those in the literature (for samples with higher molecular weight).<sup>24,25</sup> From the absence of significant Raman intensity in the ca. 1200–1260 cm<sup>-1</sup> region in the H<sub>2</sub>O sample, which appears here in the spectra of the random coil and  $\beta$ -sheet forms,<sup>24</sup> together with other differences, we can be confident that our sample-cooling procedure was effective in maintaining the  $\alpha$ -helical structure and preventing crosslinking. The ROA spectra (Fig. 2) look very different to those of L-lysine (Fig. 1), which suggests that most of the ROA intensity originates in vibrations of the peptide backbone rather than sidegroups (or that the L-lysine side-



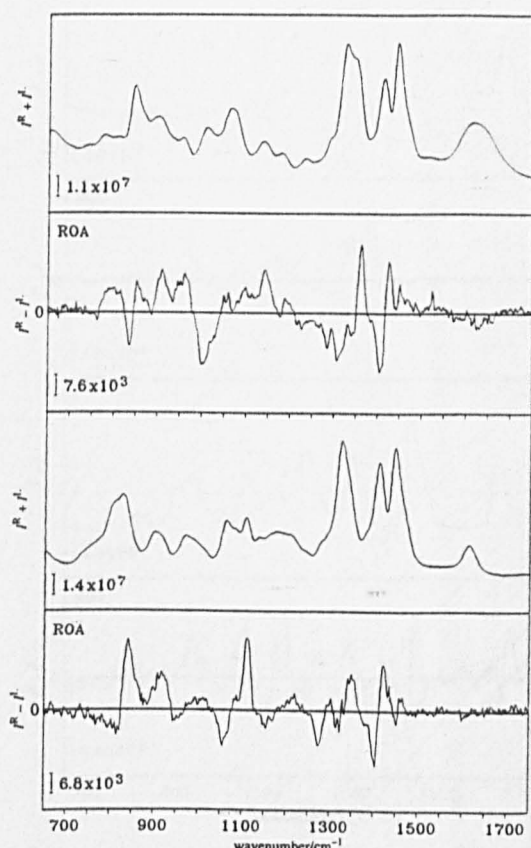


Fig. 1 Backscattered Raman ( $I^R + I^L$ ) and ROA ( $I^R - I^L$ ) spectra of L-lysine in  $H_2O$  (top half) and in  $D_2O$  (bottom half)

group conformations are very different from that of the amino acid).

Associated with the Raman band at  $ca. 943\text{ cm}^{-1}$  in  $H_2O$ , previously assigned to backbone and sidechain skeletal  $C_\alpha-C$  stretch modes of the  $\alpha$ -helix structure,<sup>3,26</sup> is a large positive ROA band peaking at  $ca. 931\text{ cm}^{-1}$  which has previously been noted as characteristic of highly  $\alpha$ -helical proteins such as bovine serum albumin and insulin,<sup>17</sup> but not of  $\beta$ -sheet proteins.<sup>18</sup> The negative ROA band at  $ca. 1000\text{ cm}^{-1}$ , which is not so prominent in  $\alpha$ -helical proteins, might also be associated with  $\alpha$ -helix modes or, since L-lysine shows negative ROA intensity here, with the sidegroups (or perhaps both helix and sidegroups). It is also possible that the positive ROA band at  $ca. 1046\text{ cm}^{-1}$  might be associated with the sidegroups, since again it does not appear to be characteristic of highly  $\alpha$ -helical proteins. However, the ROA couplet centred at  $ca. 1103\text{ cm}^{-1}$ , negative at low wavenumber and positive at high, appears to be a characteristic  $\alpha$ -helix signature since it is prominent in  $\alpha$ -helical proteins.<sup>17</sup> Although the overall ROA band pattern of  $\alpha$ -helical poly(L-lysine) in  $D_2O$  in the range  $ca. 900\text{--}1170\text{ cm}^{-1}$  is generally similar to that in  $H_2O$ , which is expected for backbone skeletal stretch modes, the ROA band intensities are much weaker. Some of the changes in relative intensity and wavenumber on deuteration can be attributed to increased involvement of N—D relative to N—H deformations in the normal modes at lower wavenumber; but this does not appear to account for such a large drop in the ROA intensities the reasons for which remain obscure.

The general appearance of the ROA band structure in the extended amide III region in  $H_2O$ , namely negative intensity

at lower wavenumber and positive at higher, is observed in most other polypeptides and all proteins studied to date. However, there can be considerable differences in the detailed structure which can be highly informative with regard to secondary structure, together with loops and turns.<sup>15–20</sup> The sharp positive ROA peak at  $ca. 1300\text{ cm}^{-1}$ , together with a smaller negative peak at  $ca. 1265\text{ cm}^{-1}$ , is also shown by  $\alpha$ -helical proteins<sup>16,17,19,20</sup> and appears to be a characteristic signature of  $\alpha$ -helix. The strong sharp positive ROA band at  $ca. 1340\text{ cm}^{-1}$  is particularly interesting since a similar band appears at this wavenumber in the ROA spectrum of bovine serum albumin (BSA) which has been ascribed to rigid loop structure with a particular type of local order.<sup>15–17</sup> As in BSA, this band vanishes in the  $D_2O$  solution, which confirms its assignment to amide III-type coupled N—H and  $C_\alpha$ —H deformations rather than  $CH_2$  deformations from the lysine sidechains. Overall the general appearance of the ROA in the extended amide III region in the  $D_2O$  solution looks similar to that of fully deuteriated lysozyme,<sup>19</sup> being characteristic of the complete loss of the N—H deformations from the corresponding normal modes everywhere in the structure which are now dominated by local  $C_\alpha$ —H deformations.<sup>19,35</sup> ROA therefore provides convincing evidence that poly(L-lysine) in its supposed standard  $\alpha$ -helical conformation (as evidenced by the appearance of the conventional Raman spectrum) in fact consists of stretches of  $\alpha$ -helix connected by rigid loop sections with the same type of local order as many of the loops in BSA. (In our earlier preliminary study, we could not be sure that this BSA-type loop structure was primarily associated with the standard  $\alpha$ -helical material because the appearance of the conventional Raman spectrum was much more unordered-like than  $\alpha$ -helix-like.<sup>17</sup>)

Evidence is accumulating that the local order associated with the loop structure responsible for the  $ca. 1340\text{ cm}^{-1}$  positive ROA band corresponds to that of  $3_{10}$ -helix,<sup>20</sup> sections of which are known to often occur at the ends of stretches of  $\alpha$ -helix.<sup>36</sup> This idea is supported by conventional FTIR studies of  $\alpha$ -helical poly(L-lysine)<sup>37</sup> in deuteriated aqueous solution which reported a strong band at  $1638\text{ cm}^{-1}$ , significantly lower than that characteristic of  $\alpha$ -helical structure in proteins but close to the wavenumber later assigned to  $3_{10}$ -helix in proteins.<sup>38</sup> Furthermore, additional FTIR amide I components close to this wavenumber were detected when the molecular weight (from viscosity) was increased from  $ca. 3300$  to  $24\,000$ , the latter being close to that employed here, which were suggested to arise 'from helical structures of slightly different hydrogen-bonding characteristics, possibly due to disruption of the hydrogen bonding in regions of the chain which become folded or twisted as the chain becomes longer'. Indeed, we found that the ROA spectrum of an  $\alpha$ -helical sample of poly(L-lysine) of molecular weight  $ca. 4000$  (not shown) displayed significantly less intensity in the band at  $ca. 1340\text{ cm}^{-1}$ , assigned to  $3_{10}$ -helical loops, and appeared to be a superposition of the ROA spectra of  $\alpha$ -helical and unordered samples. Conversely, an  $\alpha$ -helical sample of molecular weight  $ca. 268\,000$  shows significantly more intensity in the  $ca. 1340\text{ cm}^{-1}$  ROA band (*vide infra*).

It is a little surprising that the strong sharp Raman band at  $ca. 1444\text{ cm}^{-1}$  in poly(L-lysine), which originates in the side-group  $CH_2$  scissors deformation, shows virtually no ROA, in view of the fact that L-lysine itself shows large ROA in this mode. However, a couplet centred at  $ca. 1443\text{ cm}^{-1}$ , negative at low wavenumber and positive at high, appears in the ROA spectrum of the polypeptide in  $D_2O$  and can be assigned to coupling of the  $CH_2$  scissors deformation with the amide II mode, which has shifted to lower wavenumber and now overlaps with the  $CH_2$  scissors deformation band. Although the amide II itself usually shows little ROA in polypeptides and proteins, the shifted band in  $D_2O$ , called the amide II', almost

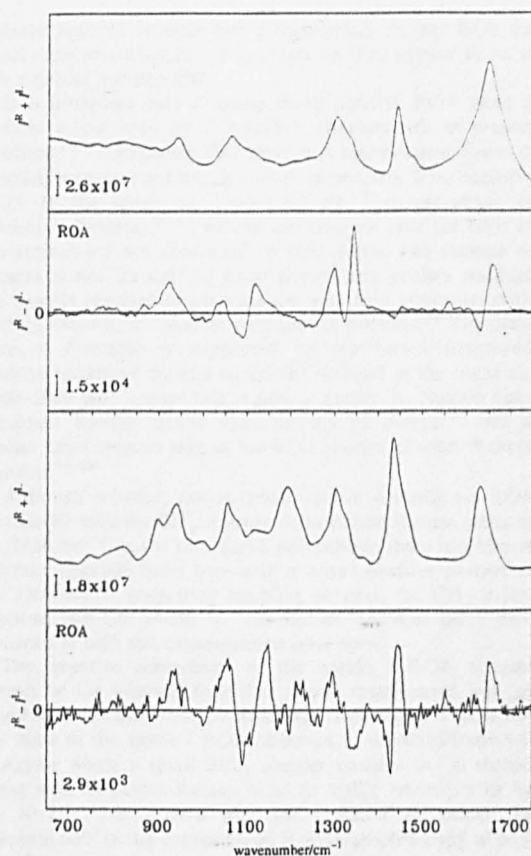


Fig. 2 Backscattered Raman and ROA spectra of  $\alpha$ -helical poly(L-lysine) ( $M_w = 26\,000$ ) in NaOH solution at pH 11.0 (top half) and in NaOD solution at pD 11.0 (bottom half) both at ca. 2°C

invariably shows significant ROA band structure due to coupling of the associated N—D deformations with sidegroup  $\text{CH}_2$  or  $\text{CH}_3$  deformations.<sup>19,35</sup>

Unlike VCD in the amide I region, which is very sensitive to different types of polypeptide and protein secondary structure,<sup>7</sup> ROA in the amide I region is not particularly informative. Both  $\alpha$ -helix and  $\beta$ -sheet proteins show amide I couplets, negative at low wavenumber and positive at high:<sup>16–18</sup> this couplet is usually conservative in  $\beta$ -sheet proteins, but has more intensity in the positive band at higher wavenumber in  $\alpha$ -helix proteins with a crossover point ca.  $10\text{ cm}^{-1}$  lower.<sup>18</sup> It is reassuring that the large positively biased amide I ROA couplet with a crossover point at ca.  $1642\text{ cm}^{-1}$  shown by  $\alpha$ -helical poly(L-lysine) is similar to that seen in  $\alpha$ -helical proteins. In  $\text{D}_2\text{O}$  solution, the amide I Raman band peak has shifted by ca.  $16\text{ cm}^{-1}$  to lower wavenumber due to the loss of the small contribution of the N—H in-plane bend to the normal mode on deuteration.<sup>2</sup> An ROA couplet similar to that in  $\text{H}_2\text{O}$  still appears, but with greatly reduced intensity as in the backbone skeletal stretch region. (In highly  $\alpha$ -helical proteins such as BSA the positively biased amide I couplet persists in  $\text{D}_2\text{O}$  with similar intensity to that in  $\text{H}_2\text{O}$  because most of the  $\alpha$ -helical structure remains unexchanged.<sup>15,16</sup>)

#### Unordered poly(L-lysine)

Again the conventional Raman spectra of the  $M_w = 26\,000$  sample of unordered poly(L-lysine) in  $\text{H}_2\text{O}$  and  $\text{D}_2\text{O}$  shown in

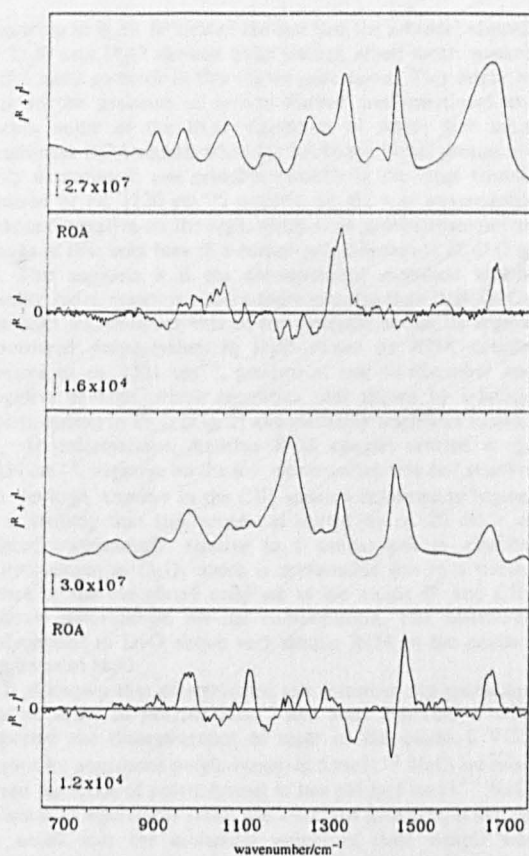


Fig. 3 Backscattered Raman and ROA spectra of unordered poly(L-lysine) ( $M_w = 26\,000$ ) in glycine buffer at pH 3.0 (top half) and in deuterated glycine buffer at pD 3.0 (bottom half), both at ca. 20°C

Fig. 3 are virtually identical to those in the literature.<sup>24,25</sup> The ROA spectra in  $\text{H}_2\text{O}$  are very similar to those obtained previously,<sup>12,17</sup> but are of better quality. It is interesting that the general appearance of these ROA spectra is quite different from those of unfolded proteins: in  $\text{H}_2\text{O}$  the latter are usually dominated by a broad relatively unstructured conservative couplet in the extended amide III region with little ROA band structure elsewhere,<sup>16,19</sup> whereas unordered poly(L-lysine) shows significant ROA structure throughout the spectrum with a positively biased, highly structured, amide III couplet. As discussed below, these observations concur with current thinking on the unordered state of homopolypeptides recent studies of which,<sup>21</sup> including VCD,<sup>39–41</sup> have provided support for the original proposal by Tiffany and Krimm<sup>42</sup> that they actually contain locally ordered stretches of left-handed helix with 2.5–3 residues per turn, possibly similar to that of poly(L-proline) II. However, in addition, the ROA data indicate that the stretches of left-handed helix are interspersed with several other well defined structures.

Thus, in place of the ROA couplet in  $\alpha$ -helical poly(L-lysine) with positive and negative peaks at ca.  $931$  and  $1000\text{ cm}^{-1}$  which appears to be a characteristic  $\alpha$ -helix signature (at least the positive peak at lower wavenumber), unordered poly(L-lysine) shows a very similar ROA couplet in this region, but with opposite sign which might originate in backbone skeletal stretch modes of left-handed helical structures. Since there is no sign of any positive ROA band at ca.  $1340\text{ cm}^{-1}$  we conclude that the regions connecting the putative stretches of left-

handed helix in random coil poly(L-lysine) do not have the local order characteristic of  $3_{10}$ -helix, as they appear to do in the  $\alpha$ -helical polypeptide.

It is intriguing that a strong sharp positive ROA band is present at  $ca. 1314\text{ cm}^{-1}$  which is characteristic of  $\beta$ -sheet proteins,<sup>16,18</sup> indicating that there is a significant amount of  $\beta$ -strand structure present; however, since there is no negative ROA in the range  $ca. 1340\text{--}1380\text{ cm}^{-1}$  characteristic of  $\beta$ -turns in proteins,<sup>16,18</sup> we can also suggest that the ends of the strands are not connected by tight turns. The absence of  $\beta$ -turns is not unexpected since glycine and proline residues are usually required to generate the tight turn structures connecting the ends of adjacent  $\beta$ -strands in proteins.<sup>36</sup> The presence of  $\beta$ -strands is supported by the broad structured positive backbone skeletal stretch ROA band in the range  $ca. 1000\text{--}1060\text{ cm}^{-1}$ , since this region is known to contain conventional Raman bands characteristic of  $\beta$ -sheet,<sup>3</sup> and a similar band appears here in the ROA spectra of some  $\beta$ -sheet proteins.<sup>16,18</sup>

Although  $\alpha$ -helical poly(L-lysine) shows virtually no ROA associated with the  $\text{CH}_2$  scissors deformation Raman band at  $ca. 1444\text{ cm}^{-1}$ , in the unordered polypeptide there is a hint of a broad negative band here with a broad positive partner at  $ca. 1550\text{ cm}^{-1}$ , suggesting coupling between the  $\text{CH}_2$  deformations and the amide II: this feature becomes more pronounced at high salt concentration (*vide infra*).

The negative component of the amide I ROA couplet shown by the  $\alpha$ -helical form has almost disappeared but the positive component remains relatively unchanged. This is not the same as the amide I ROA observed in unfolded (reduced) lysozyme where a small ROA couplet remains but is shifted along with its parent Raman band to higher wavenumber by  $ca. 10\text{ cm}^{-1}$ , in accord with the amide I assignment of 'random coil' in the conventional Raman spectroscopy of proteins.<sup>3</sup>

It is interesting that a sharp amide III ROA couplet, with a weaker negative component at  $ca. 1265\text{ cm}^{-1}$  and a stronger positive component at  $ca. 1300\text{ cm}^{-1}$ , is clearly present on the low wavenumber side of the strong positive  $\beta$ -strand ROA band at  $ca. 1314\text{ cm}^{-1}$ , since this signature also appears in the high-pH sample of poly(L-lysine) where it has been assigned to  $\alpha$ -helix vibrations (*vide supra*). If this assignment is correct, then sections of  $\alpha$ -helix are also present in the low-pH sample of poly(L-lysine). At first sight it might be thought that this is incompatible with the suggested presence of left-handed helix, because an amide III couplet at similar wavenumber but opposite sign might have been expected, which would have tended to cancel that from the  $\alpha$ -helix. However, interchain hydrogen bonds are absent from the poly(L-proline) II helix,<sup>36,39</sup> and it has been suggested that interchain hydrogen bonds are also absent from residues taking up this conformation in proteins (the conformation being stabilized by hydrogen bonds to either water or sidechains).<sup>21,43</sup> Interchain hydrogen bonding of N—H groups has a significant influence on the contribution of N—H deformation coordinates to normal modes in the extended amide III region, so we would not expect left-handed helix to show simple mirror-image versions of the  $\alpha$ -helix ROA signatures here. However, since N—H deformations contribute much less to normal modes in the backbone skeletal stretch region, the absence of the ROA couplet, negative at lower wavenumber and positive at higher, centred at  $ca. 1103\text{ cm}^{-1}$  in  $\alpha$ -helical poly(L-lysine) might well be due to cancellation by a similar ROA couplet of opposite sign from the left-handed helix (in fact a small couplet of opposite sign can be seen here which might be the residual left-handed helix couplet).

It is perplexing that the ROA spectrum of unordered poly(L-lysine) in  $\text{D}_2\text{O}$  in the backbone skeletal stretch region, and also on the low wavenumber side of the extended amide III region, shows little relation to the corresponding ROA

spectrum in  $\text{H}_2\text{O}$ , in view of the fact that the  $\alpha$ -helical samples in  $\text{D}_2\text{O}$  and  $\text{H}_2\text{O}$  showed quite similar, albeit much weaker, ROA band patterns in this region (*vide supra*). This might be due to the presence of several distinct conformational elements, some of the ROA signatures of which give more prominent ROA signals when the backbone N—H groups are fully deuterated: one possible example is the large couplet centred at  $ca. 1130\text{ cm}^{-1}$ , positive on the low wavenumber side and negative on the high, which is an approximate mirror image of that seen here in  $\alpha$ -helical poly(L-lysine) in  $\text{D}_2\text{O}$  (Fig. 2). This suggests it is the corresponding signature of left-handed helix, appearing much more strongly than in  $\text{H}_2\text{O}$ . On the high wavenumber side of the extended amide III region, unordered poly(L-lysine) in  $\text{D}_2\text{O}$  shows an ROA couplet centred at  $ca. 1331\text{ cm}^{-1}$ , positive at low wavenumber and negative at high, which resembles that shown by  $\alpha$ -helical poly(L-lysine) in  $\text{D}_2\text{O}$  (Fig. 2) and probably originates in local  $\text{C}_\alpha\text{—H}$  deformations. Another ROA couplet centred at  $ca. 1454\text{ cm}^{-1}$ , negative on the low wavenumber side and positive on the high, appears in the  $\text{CH}_2$  scissors deformation region: it is striking that this couplet is shifted by  $ca. 20\text{ cm}^{-1}$  to higher wavenumber, relative to a similar one in  $\alpha$ -helical poly(L-lysine) in  $\text{D}_2\text{O}$ , which is presumably due to a dependence of the associated coupling of the amide II' and  $\text{CH}_2$  scissors deformation on the conformation. The unordered polypeptide in  $\text{D}_2\text{O}$  shows very similar ROA in the amide I region as in  $\text{H}_2\text{O}$ .

It is known that concentrated salt solutions can encourage loss of order in polypeptides,<sup>44</sup> and since Paterlini *et al.*<sup>39</sup> reported the disappearance of most of the amide I VCD shown by unordered poly(L-lysine) in  $5\text{ mol l}^{-1}\text{ NaCl}$  we measured the ROA of poly(L-lysine) at low pH in  $5\text{ mol l}^{-1}\text{ NaCl}$  to see if an equivalent result was obtained (however, it should be noted that the molecular weight of their sample was 525 000 whereas ours was only 26 000). The corresponding Raman and ROA spectra measured at  $ca. 20^\circ\text{C}$  are shown in the top half of Fig. 4. Unlike the resulting VCD spectrum, the ROA spectrum is still highly structured, which implies that considerable ordered structure remains in the polypeptide: in particular, the negative ROA band at  $ca. 926\text{ cm}^{-1}$  in the unordered sample (Fig. 3) ascribed to left-handed helical structure is much weaker and most of the ROA features in the backbone skeletal stretch, amide III and amide I regions ascribed to  $\alpha$ -helix from the high pH sample (Fig. 2) have increased; whereas the sharp positive peak at  $ca. 1314\text{ cm}^{-1}$ , assigned to  $\beta$ -strand, is unchanged. Notice, in particular, that a negative-positive ROA couplet centred at  $ca. 1103\text{ cm}^{-1}$ , characteristic of  $\alpha$ -helix, has now appeared (with the negative and positive peaks shifted slightly apart). This supports the suggestion above that the absence of this feature in the unordered sample is due to cancellation by a similar signal of opposite sign from the left-handed helix. Furthermore, the broad couplet, negative at  $ca. 1448\text{ cm}^{-1}$  and positive at  $ca. 1554\text{ cm}^{-1}$ , which is just perceptible in the unordered sample (Fig. 3) but is not apparent in the  $\alpha$ -helical sample, ascribed to coupling between sidegroup  $\text{CH}_2$  deformations and the backbone amide II, is now much more pronounced. As can be seen from the ROA spectrum in the bottom half of Fig. 4, on heating the sample to  $ca. 50^\circ\text{C}$  this trend towards increased intensity of  $\alpha$ -helix ROA signatures at the expense of those of left-handed helix is reinforced. Hence, ROA shows that both left-handed helix, right-handed  $\alpha$ -helix and  $\beta$ -strand structures are represented in unordered poly(L-lysine), with the left-handed helix giving way to right-handed  $\alpha$ -helix under destabilizing conditions. This conclusion brings to mind the study by Drake *et al.*<sup>45</sup> of the temperature dependence of the electronic UV CD of poly(L-lysine) solutions over the range  $-100$  to  $+80^\circ\text{C}$ . This suggested that conformers containing left-handed helical structure become dominant at low temperature;<sup>21</sup> however, unlike the present ROA study, these UV



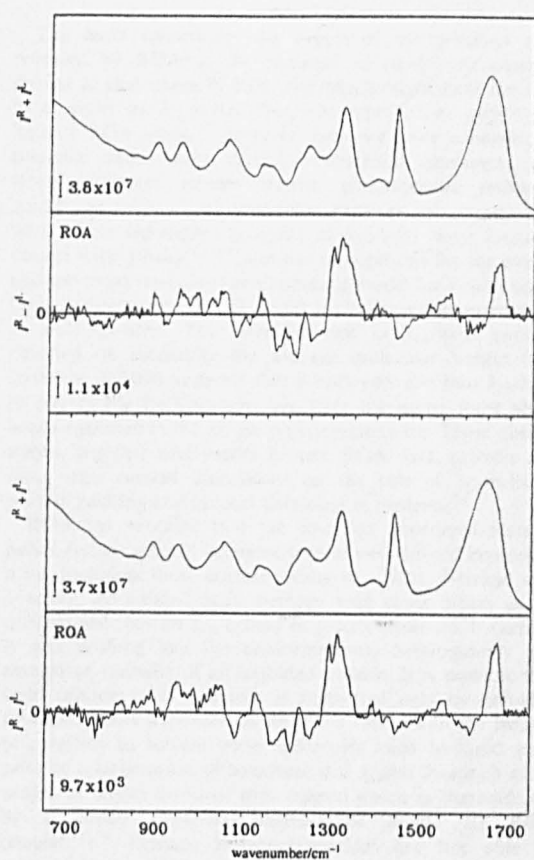


Fig. 4 Backscattered Raman and ROA spectra of unordered poly(L-lysine) ( $\bar{M}_w = 26\,000$ ) in glycine buffer at pH 3.0 with the addition of 5 mol l<sup>-1</sup> NaCl at ca. 20 °C (top half) and ca. 50 °C (bottom half)

CD studies were unable to provide much insight into the nature of the high-temperature state.

#### Higher-molecular-weight poly(L-lysine)

In view of the possibility that the predominance of particular conformational features might depend on the chain length of individual polymers, we measured the ROA of  $\alpha$ -helical and unordered poly(L-lysine) using a sample with  $\bar{M}_w = 268\,000$ , this being roughly the highest value for which ROA measurements are possible (on account of lower solubility, increased tendency to gel formation *etc.*). The resulting spectra are shown in Fig. 5.

The ROA spectrum of the supposedly  $\alpha$ -helical form of the higher-molecular-weight poly(L-lysine), shown in the top half of Fig. 5, displays some remarkable changes relative to that of the lower molecular weight sample in Fig. 2. The positive ca. 1340 cm<sup>-1</sup> band, assigned to loop structure with local order corresponding to 3<sub>10</sub>-helix, has grown to ca. 1.5 times its previous intensity, with the intensities of the amide I and amide III  $\alpha$ -helix bands, together with that of the  $\alpha$ -helix couplet centred at ca. 1103 cm<sup>-1</sup>, falling to ca. half their previous values, suggesting an increase in the amount of rigid 3<sub>10</sub>-helical loop structure at the expense of  $\alpha$ -helix. Additionally, the sharp positive  $\alpha$ -helix ROA band at ca. 931 cm<sup>-1</sup> has been replaced by broad weak ROA intensity peaking at ca. 838 cm<sup>-1</sup> which might be associated with ROA intrinsic to 3<sub>10</sub>-helix, and the negative and positive ROA bands at ca. 1000 and 1046 cm<sup>-1</sup> have disappeared. This large increase in 3<sub>10</sub>-

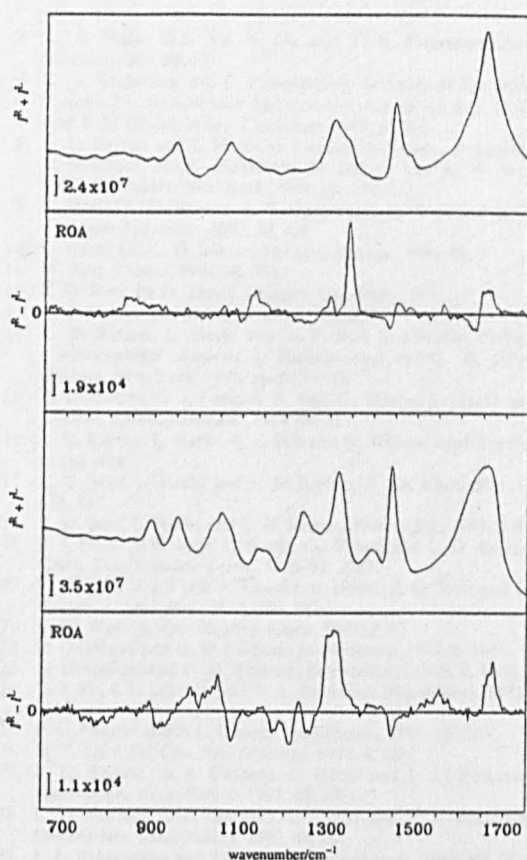


Fig. 5 Backscattered Raman and ROA spectra poly(L-lysine) ( $\bar{M}_w = 268\,000$ ),  $\alpha$ -helical in NaOH solution at pH 11.0 and ca. 2 °C (top half) and unordered in glycine buffer at pH 3.0 and ca. 20 °C (bottom half)

helix content on increasing the average molecular weight from 24 000 to 268 000 might be a measure of the greater tangling from the many more sharp turns which the more regular  $\alpha$ -helix cannot support.

The ROA spectrum of the unordered higher molecular weight sample of poly(L-lysine), shown in the bottom half of Fig. 5, displays a general reinforcement and sharpening of the features seen in that of the lower-molecular-weight sample in Fig. 3. One difference, however, is a large increase in the negative band at ca. 1075 cm<sup>-1</sup>, which now makes up a prominent couplet with the positive band peaking at ca. 1048 cm<sup>-1</sup>. A similar couplet appears in the ROA spectrum of the lower-molecular-weight unordered sample at high salt concentration shown in Fig. 4. This feature could be associated with the right-handed twist characteristic of the strands in  $\beta$ -sheet which is equivalent to a left-handed helical conformation of the peptide backbone,<sup>46</sup> for a similar couplet appears in proteins such as ribonuclease A<sup>13</sup>, which contain highly twisted double-stranded  $\beta$ -sheet but not in proteins such as concanavalin A<sup>16</sup> which contain relatively flat multi-stranded  $\beta$ -sheet in a  $\beta$ -barrel.

#### Concluding remarks

On account of its ability to monitor both extended secondary structure and locally ordered loop structure, ROA has provided a fresh perspective on the supposedly  $\alpha$ -helical and random coil states of poly(L-lysine) in aqueous solution.



The most interesting new aspect of the  $\alpha$ -helical state, revealed by ROA, is the presence of rigid loop structure similar to that found in BSA, and which might have the same local order as  $3_{10}$ -helix. This was reported in earlier preliminary ROA work:<sup>17</sup> however, here, we have confirmed its presence under more carefully controlled conditions. Our work reinforces recent electron paramagnetic resonance studies on small  $C_{\alpha}$  disubstituted peptides in aqueous solution, where significant amounts of  $3_{10}$ -helix were found to coexist with  $\alpha$ -helix,<sup>47-50</sup> and has implications for the continued use of poly(L-lysine) as a standard model for  $\alpha$ -helix signatures in electronic UV CD, which is challenged by mixtures of  $\alpha$ - and  $3_{10}$ -helix. The large increase in  $3_{10}$ -helix content, detected on increasing the average molecular weight from 26 000 to 268 000, suggests that  $\alpha$ -helix converts into  $3_{10}$ -helix to enable the backbone to negotiate the many more sharp bends expected in the longer polypeptide chain. These observations, together with recent protein ROA data, provide new input into current discussions on the role of  $3_{10}$ -helix in protein packing and in local unfolding in proteins.<sup>51</sup>

ROA has revealed that the so-called unordered state of poly(L-lysine) actually contains several well defined conformations, including those corresponding to  $\alpha$ -helix,  $\beta$ -strand and, possibly, left-handed helix, perhaps with some others as yet unidentified; but no  $3_{10}$ -helical or  $\beta$ -turn structure. It certainly has nothing like the conformational heterogeneity and associated mobility of an unfolded protein. It is possible that true 'random coil' structure is supported only by unfolded proteins, where exposure of all of the residues in the heteropolypeptide to solvent water molecules leads to rapid sampling of a large range of backbone  $\phi, \psi$  angles (but with some ranges of angles favoured over others) which is characterized by a broad relatively unstructured amide III ROA couplet.<sup>16,19</sup> Perhaps homopolypeptides are not able to support this type of dynamic disorder?

The close similarity of a number of features in the ROA spectrum of unordered poly(L-lysine) with those found in proteins containing highly twisted  $\beta$ -sheet<sup>52</sup> prompts the suggestion that the putative left-handed helix is perhaps more akin to that of right-twisted  $\beta$ -strand<sup>46</sup> than to that of poly(L-proline II).

This qualitative analysis of the ROA spectra of different states of poly(L-lysine) has been somewhat overwhelmed by the enormous wealth of detailed information they contain about the distinct conformational features present. The existence of several independent ROA signatures of certain types of structure suggests that it should ultimately be possible to specify with reasonable precision the actual compositions. However, the ROA spectra of each distinct conformer cannot be measured separately since it does not appear to be possible to prepare pure samples containing each in isolation, so this type of quantitative analysis will probably have to wait until the ROA spectra of model peptide conformations can be calculated reliably using *ab initio* methods.<sup>29</sup>

We thank the EPSRC and BBSRC for research grants, the EPSRC for a Senior Fellowship for L.D.B. and a Research Studentship for G.W. and Professor T. A. Keiderling for helpful discussions

## References

- 1 P. R. Carey, *Biochemical Applications of Raman and Resonance Raman Spectroscopies*, Academic Press, New York, 1982.
- 2 S. Krimm and J. Bandekar, *Adv. Protein Chem.*, 1986, 38, 181.
- 3 A. T. Tu, in *Advances in Spectroscopy, Volume 13: Spectroscopy of Biological Systems*, ed. R. J. H. Clark and R. E. Hester, Wiley, Chichester, 1986, p. 47.
- 4 L. D. Barron, *Molecular Light Scattering and Optical Activity*, Cambridge University Press, Cambridge, 1982.
- 5 M. Diem, *Modern Vibrational Spectroscopy*, Wiley, New York, 1993.
- 6 L. A. Nafie, G.-S. Yu, X. Qu and T. B. Freedman, *Faraday Discuss.*, 1994, 99, 13.
- 7 T. A. Keiderling and P. Pancoska, in *Advances in Spectroscopy, Volume 21: Biomolecular Spectroscopy, Part B*, ed. R. J. H. Clark and R. E. Hester, Wiley, Chichester, 1993, p. 267.
- 8 L. D. Barron and L. Hecht, in *Circular Dichroism: Principles and Applications*, ed. K. Nakanishi, N. Berova and R. W. Woody, VCH Publishers, New York, 1994, pp. 179-215.
- 9 L. Hecht, L. D. Barron, A. R. Gargaro, Z. Q. Wen and W. Hug, *J. Raman Spectrosc.*, 1992, 28, 401.
- 10 L. Hecht and L. D. Barron, *Faraday Discuss.*, 1994, 99, 35.
- 11 W. Hug, *Chimia*, 1994, 48, 386.
- 12 Z. Q. Wen, Ph.D. Thesis, Glasgow University, 1992.
- 13 L. D. Barron and L. Hecht, in ref. 7, p. 235.
- 14 L. D. Barron, L. Hecht and A. F. Bell, in *Circular Dichroism: Conformational Analysis of Biomolecules*, ed. G. D. Fasman, Plenum, New York, 1996, pp. 653-695.
- 15 L. D. Barron, S. J. Ford, A. F. Bell, G. Wilson, L. Hecht and A. Cooper, *Faraday Discuss.*, 1994, 99, 217.
- 16 L. D. Barron, L. Hecht, A. F. Bell and G. Wilson, *Appl. Spectrosc.*, in the press.
- 17 Z. Q. Wen, L. Hecht and L. D. Barron, *J. Am. Chem. Soc.*, 1994, 116, 443.
- 18 Z. Q. Wen, L. Hecht and L. D. Barron, *Protein Sci.*, 1994, 3, 435.
- 19 S. J. Ford, A. Cooper, L. Hecht, G. Wilson and L. D. Barron, *J. Chem. Soc., Faraday Trans.*, 1995, 91, 2087.
- 20 G. Wilson, S. J. Ford, A. Cooper, L. Hecht, Z. Q. Wen and L. D. Barron, *J. Mol. Biol.*, 1995, 254, 747.
- 21 R. W. Woody, *Adv. Biophys. Chem.*, 1992, 2, 37.
- 22 B. Davidson and G. D. Fasman, *Biochemistry*, 1967, 6, 1616.
- 23 N. Greenfield and G. D. Fasman, *Biochemistry*, 1969, 8, 4108.
- 24 T. J. Yu, J. L. Lippert and W. L. Peticolas, *Biopolymers*, 1973, 12, 2161.
- 25 P. C. Painter and J. L. Koenig, *Biopolymers*, 1976, 15, 229.
- 26 N. T. Yu, *CRC Crit. Rev. Biochem.*, 1977, 4, 229.
- 27 L. D. Barron, A. R. Gargaro, L. Hecht and P. L. Polavarapu, *Spectrochim. Acta, Part A*, 1991, 47, 1001.
- 28 L. D. Barron, A. R. Gargaro, L. Hecht and P. L. Polavarapu, *Spectrochim. Acta, Part A*, 1992, 48, 261.
- 29 P. L. Polavarapu and Z. Deng, *Faraday Discuss.*, 1994, 99, 151.
- 30 A. R. Gargaro, L. D. Barron and L. Hecht, *J. Raman Spectrosc.*, 1993, 24, 91.
- 31 F. R. Dollish, W. G. Fateley and F. F. Bentley, *Characteristic Raman Frequencies of Organic Compounds*, Wiley, New York, 1974.
- 32 N. B. Colthup, L. H. Daley and S. E. Wiberley, *Introduction to Infrared and Raman Spectroscopy*, Academic Press, New York, 1975.
- 33 R. C. Lord and N.-T. Yu, *J. Mol. Biol.*, 1970, 50, 509.
- 34 M. Diem, O. Lee and G. M. Roberts, *J. Phys. Chem.*, 1992, 96, 548.
- 35 S. J. Ford, Z. Q. Wen, L. Hecht and L. D. Barron, *Biopolymers*, 1994, 34, 303.
- 36 T. E. Creighton, *Proteins*, Freeman, New York, 1993.
- 37 M. Jackson, P. I. Harris and D. Chapman, *Biochim. Biophys. Acta*, 1989, 998, 75.
- 38 S. J. Prestrelski, D. M. Blyler and M. P. Thompson, *Int. J. Peptide Protein Res.*, 1991, 37, 508.
- 39 M. G. Paterlini, T. B. Freedman and L. A. Nafie, *Biopolymers*, 1986, 25, 1751.
- 40 R. K. Dukor and T. A. Keiderling, *Biopolymers*, 1991, 31, 1747.
- 41 S. S. Birke, I. Agbaje and M. Diem, *Biochemistry*, 1992, 31, 450.
- 42 M. L. Tiffany and S. Krimm, *Biopolymers*, 1968, 6, 1379.
- 43 A. A. Adzhubei, F. Eisenmenger, V. G. Tumanyan, M. Zinke, S. Brodzinski and N. G. Esipova, *Biochem. Biophys. Res. Commun.*, 1987, 146, 934.
- 44 M. L. Tiffany and S. Krimm, *Biopolymers*, 1969, 8, 347.
- 45 A. F. Drake, G. Siligardi and W. A. Gibbons, *Biophys. Chem.*, 1988, 31, 143.
- 46 P. H. Maccallum, R. Poet and E. J. Milner-White, *J. Mol. Biol.*, 1995, 248, 374.
- 47 S. M. Mick, G. V. Martinez, W. R. Fiori, A. P. Todd and G. L. Millhauser, *Nature (London)*, 1992, 359, 653.
- 48 W. R. Fiori, S. M. Mick and G. L. Millhauser, *Biochemistry*, 1993, 32, 11957.
- 49 W. R. Fiori, K. M. Lundberg and G. L. Millhauser, *Nature Struct. Biol.*, 1994, 1, 374.
- 50 G. L. Millhauser, *Biochemistry*, 1995, 34, 3873.
- 51 G. Wilson, L. Hecht and L. D. Barron, to be published.

Paper S/07705B; Received 27th November, 1995